Structural and Biochemical Studies of

Oct-1 POU Domain—DNA Interactions

by

Juli Dawn Klemm

B. S., Biology
University of Utah, 1989

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the
Massachusetts Institute of Technology

July 1995

Signature of Author.....
Department of Biology

Certified by.....
Carl O. Pabo, Professor of Biophysics and Structural Biology
Thesis Supervisor

Accepted by.....
Frank Solomon, Professor of Biology
Chairman, Biology Graduate Committee

STRUCTURAL AND BIOCHEMICAL STUDIES OF OCT-1 POU DOMAIN—DNA INTERACTIONS

by
Juli Dawn Klemm

submitted to the Department of Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

This thesis describes studies of the POU domain from the transcriptional regulatory protein Oct-1. The POU domain is a bipartite DNA-binding structure found in a variety of eukaryotic transcription factors. It is comprised of two structurally independent subdomains: a POU-specific domain and a POU homeodomain are connected by a flexible polypeptide linker. We have solved the x-ray crystal structure of the Oct-1 POU domain complexed with DNA and we have performed biochemical experiments to measure cooperative interactions between the POU-specific domain and the POU homeodomain.

Chapter 1 discusses three DNA-binding motifs: POU domains, TFIIIA-type zinc fingers, and paired domains. These three motifs are each comprised of structurally independent DNA-binding modules connected by polypeptide linkers. This chapter compares the use of linked modules in DNA recognition by these three DNA-binding structures.

Chapter 2 describes the x-ray crystal structure of the Oct-1 POU domain complexed with its DNA recognition element, the octamer site. This chapter was first published as "Crystal Structure of an Oct-1 POU Domain—Octamer Complex: DNA Recognition With Tethered DNA-Binding Modules" (Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. 1994. Cell **77:** 21-32).

Chapter 3 describes experiments addressing issues of cooperativity in Oct-1 POU domain—DNA interactions. These studies reveal that individual

peptides corresponding to the Oct-1 POU-specific domain and POU homeodomain bind cooperatively to the octamer sequence. These studies have also allowed for quantitation of the effective concentration of the subdomains in the context of the intact POU domain.

Chapter 4 summarizes what has been learned about Oct-1 POU domain —DNA interactions and discusses prospects for future research regarding the POU family of transcription factors and cooperativity in other protein-DNA interactions.

Thesis supervisor: Carl O. Pabo
Title: Professor of Biophysics and Structural Biology

This thesis is dedicated to my parents,

Donald and Joan Klemm

Acknowledgments

There are many people who have made my graduate work at MIT productive and enjoyable and I would like to take this opportunity to offer my appreciation to them.

First of all, I would like to thank Carl Pabo, my advisor, for his guidance, encouragement, generosity, and many interesting discussions. Carl's sincere interest in providing the best possible environment for research and learning has made the lab a great place to be a graduate student.

I wish to thank the members of my thesis committee - Bob Sauer, Phil Sharp, and Peter Kim - for their interest and advice throughout my graduate career. I would also like to thank Tom Ellenberger for serving on my thesis defense committee.

I have sincerely enjoyed working with all of the members of the Pabo lab, past and present. I have greatly benefited from the help I have received from everyone and I certainly hope that in the future I can work with an equally talented group of scientists. In particular, Kristen Chambers, Harvey Greisman, Ernest Fraenkel and Ed Rebar have provided not only invaluable scientific advice, but have also provided lots of laughs in the lab. Mark Rould, crystallographer extraordinaire, has spent many hours teaching me crystallography and has been great fun to work with. Finally, I must thank Nikola Pavletich (now with his own lab at Sloan-Kettering) for spending so much time getting me up to speed when I joined the group, and for being such a motivating force in the lab.

My work with the Oct-1 POU domain has involved a productive collaboration with Winship Herr's lab at Cold Spring Harbor. Winship has been an inexhaustible source of enthusiasm and I have learned a great deal

from working with him. I have also enjoyed working with Reggie Aurora and Michele Cleary, both of whom taught me much about molecular biology during my trips to Cold Spring Harbor.

Outside of the lab, I've had a lot of good times with the members of my graduate class, including many dinners out, birthday parties, barbecues, and sports events. It hard to imagine that I'll find such a large group of friends when I leave MIT. A special thanks to Brenda Schulman for three enjoyable years at 12 Winter Street.

I owe a debt of gratitude to one classmate in particular, Mike, for seeing me through it all with his patience, sense of humor, understanding, and love.

Finally, I thank my parents for a lifetime of encouragement and support.

Biographical Note

Juli D. Klemm

EDUCATION

Massachusetts Institute of Technology — present
Ph.D. candidate, Department of Biology

University of Utah — Dec. 1989
Bachelor of Science, Biology
Honors at Entrance Scholarship

PROFESSIONAL EXPERIENCE

RESEARCH

M.I.T., Department of Biology — 1990 - present
Graduate Student
Supervisor: Professor Carl O. Pabo
Thesis topic: Structural and biochemical studies of Oct-1 POU domain—DNA interactions

University of Utah, Department of Biochemistry — 1988 - 1990
Undergraduate research
Supervisor: Professor Tom Alber
Fields of Study: Protein folding and protein x-ray crystallography

University of Utah, Department of Biology — 1987 - 1988
Undergraduate research
Supervisor: Professor K. G. Lark
Field of Study: Molecular analysis of the soybean plant

TEACHING

Teaching Assistant, M.I.T.
Courses: "Introductory Biology" — 1992
"Introductory Biology Lab" — 1994

PUBLICATIONS

Klemm, J. D., and Pabo, C. O. (1995). Oct-1 POU Domain—DNA Interactions: Cooperative Binding of Isolated Subdomains and Effects of Covalent Linkage. (submitted)

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal Structure of the Oct-1 POU Domain Bound to an Octamer Site: DNA Recognition with Tethered DNA-Binding Molecules. Cell 77, 21-32.

O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991). X-ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil. Science 254, 539-544.

Klemm, J. D., Wozniak, J. A., Alber, T., and Goldenberg, D. P. (1991). Correlation between Mutational Destabilization of Phage T4 Lysozyme and Increased Unfolding Rates. Biochemistry 30, 589-594.

TABLE OF CONTENTS

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

CHAPTER 1

Covalently Linked DNA-Binding Modules: POU Domains, Zinc Fingers and Paired Domains

Introduction

The study of eukaryotic transcriptional regulation has made a tremendous contribution to our understanding of key biological processes such as cell development, cell growth, and cellular responses to external signals. Many of these studies have focused on transcription factors, which are the proteins that specifically bind to the cis-acting regions of genes to activate or repress the synthesis of mRNA. Transcription factors are generally comprised of discrete functional domains. For example, a typical transcriptional activator will contain a DNA-binding domain, which makes sequence-specific contacts with the DNA; an activation domain, which interacts with other components of the transcriptional apparatus; and in many cases, a multimerization domain, which is used to form dimers or higher-order multimers (Tjian and Maniatis, 1994). "Domain-swap" experiments have revealed that the functional domains from different transcription factors can be recombined to form new proteins with novel functions (see Ptashne, 1992). Such experiments demonstrate the modular nature of transcription factors and the ease with which their activation domains and DNA binding domains can be combined (for review, see Frankel and Kim, 1991).

The modular nature of eukaryotic transcription factors allows their composite units to be studied individually. This feature has been useful for structural studies where smaller, globular domains are considerably easier to study than the entire parent molecule. To date, it has not been possible to obtain structural information about activation domains, and some of these may be unstructured in the absence of interacting proteins. Thus, structural studies of eukaryotic transcription factors have focused almost exclusively on DNA-binding domains and oligomerization domains. These studies have

been extremely fruitful: the x-ray crystal structures of more than 30 different transcription factor DNA-binding domain—DNA complexes have been solved in the past decade and have provided a wealth of information about the structural basis of specificity in transcription factor-DNA interactions (Pabo and Sauer, 1992; Ades, 1995).

The POU domain is a DNA-binding motif present in a family of eukaryotic transcription factors involved in various aspects of transcriptional regulation. This motif contains two structurally independent domains - a POU-specific domain and a POU homeodomain - that cooperate functionally as a DNA-binding unit. In order to understand how this important class of transcriptional regulatory proteins binds DNA, we have solved the x-ray crystal structure of the POU domain from the protein Oct-1 bound to its recognition element, the octamer site. This structure has provided a wealth of information about POU domain—DNA interactions and, in addition, prompted us to pursue biochemical experiments addressing subdomain cooperativity in DNA recognition by the POU domain.

This introductory chapter begins with a brief background on the POU family of transcription factors and summarizes our structural and biochemical results with the Oct-1 POU domain (which will be described in detail in Chapters 2 and 3). These results have given us a physical and energetic description of how the POU domain uses its covalently linked DNA-binding modules to interact with DNA, and provide a framework for thinking about how other DNA-binding domains comprised of covalently-linked modules recognize DNA. Thus, in the second part of the chapter, two other DNA-binding motifs, TFIIIA-type zinc fingers and paired domains, will be discussed in terms of how they use covalently-linked DNA binding domains in recognition. All three families of DNA-binding domains (POU

domains, zinc fingers, and paired domains) will then be compared on the basis of: 1) how their individual subdomains contribute to DNA binding; and 2) the characteristics of the linkers connecting the subdomains. Covalently linked DNA-binding modules provide an interesting level of complexity for understanding protein-DNA recognition, and the unique features of these motifs are discussed in the final sections.

The POU domain

The POU domain has been found in ~30 transcription factors from various metazoan organisms; its name is derived from the first four proteins identified containing this motif (Pit-1, Oct-1 and -2, and Unc-86) (Herr et al., 1988). POU domain transcription factors have been shown to regulate many different types of genes. For example, Unc-86, a *C. elegans* POU domain protein, is required in several neuroblast lineages for daughter cells to become different from the mother cells (Finney et al., 1988). Cf1a, a *Drosophila* POU domain protein, is involved in the regulation of DOPA decarboxylase gene expression (Johnson and Hirsh, 1990). Oct-1, one of the best characterized POU domain proteins, is ubiquitously expressed and has been implicated in variety of regulatory functions, including: 1) histone H2B gene expression; 2) small nuclear RNA gene expression; 3) immunoglobulin gene expression; 4) herpes simplex virus immediate early gene expression; 5) mouse mammary tumor virus gene expression; and 6) stimulation of adenovirus replication (reviewed by Rosenfeld, 1991; Ruvkin and Finney, 1991; Verrijzer and van der Vliet, 1993). Thus, POU domain proteins play important roles in development and differentiation, as well as in general housekeeping functions.

The first POU domain proteins were identified as homeobox-containing proteins, and subsequent sequence comparison revealed a 75 amino acid region of strong homology N-terminal to the homeodomain, termed the POU-specific domain (Herr et al., 1988). The POU-specific domain and the POU homeodomain are connected by a linker region that is highly variable in length and sequence among POU domain proteins. It was known from studies of other homeobox proteins that the homeodomain was a sequence-specific DNA-binding domain; however, it was not known what role the highly conserved POU-specific domain played in this family of transcription factors.

Biochemical analysis of the POU domain revealed that both the POU-specific domain and the POU homeodomain are required for high affinity DNA binding by this family of transcription factors. Sturm and Herr (1988) showed that disrupting either Oct-1 POU subdomain by inserting three alanine residues eliminated DNA binding by this protein, demonstrating that both domains are required for recognition. Similarly, experiments by Ingraham et al. (1990) with Pit-1 showed that the protein with its POU-specific domain removed retained some DNA binding activity but the protein with its POU homeodomain removed had no detectable DNA binding activity. These results, showing that the POU-specific domain is required for optimal DNA binding by POU domain proteins, were somewhat surprising in light of the fact that the DNA-binding activity of other homeobox proteins had been shown to reside exclusively with the homeodomain. Thus, it was clear that the POU family of homeodomain proteins had properties distinct from other well-characterized homeodomain proteins.

Although these studies confirmed that the POU-specific domain (in addition to the homeodomain) plays an important role in DNA recognition

by POU domain proteins, it had not been determined whether the POU-specific domain contributes to DNA binding by directly contacting the DNA or by indirectly affecting the contacts made by the POU homeodomain. Comparison of the footprinting patterns of the Oct-1 POU homeodomain and the intact Oct-1 POU domain showed that deletion of the POU-specific domain resulted in a specific loss of protection at the "left half" of the binding site, strongly suggesting that the POU-specific domain is responsible for those contacts (Verrijzer et al., 1990; Kristie and Sharp, 1990). Direct evidence that the POU-specific domain contacts the DNA was provided by crosslinking experiments by Aurora and Herr (1992). Their results demonstrated that both the POU-specific domain and the POU homeodomain can be crosslinked to the major groove of BrdU-substituted DNA, confirming that both subdomains are in direct contact with the DNA.

Experiments with the individual POU-specific domain and POU homeodomain demonstrated that these two subdomains are structurally and functionally distinct. Botfield et al. (1992) performed extensive biophysical studies with the POU-specific domain and POU homeodomain polypeptides from the Oct-2 protein. Using a wide variety of experimental techniques, they demonstrated that: 1) the POU-specific domain and the POU homeodomain are proteolytically stable subdomains, but the linker (in the context of the intact POU domain) is proteolytically sensitive; 2) each subdomain exhibits cooperative folding/unfolding transitions, as assayed by circular dichroism; and 3) ^{1}H NMR spectra of the isolated subdomains can be added together to give a spectrum nearly identical to the spectrum of the intact POU domain. These data strongly suggested that the POU domain is comprised of two independent domains tethered by a flexible linker. Verrijzer et al. (1992) performed experiments with the isolated Oct-1 POU subdomains with the

goal of dissecting the contribution of each subdomain to the overall specificity of the intact POU domain. Importantly, they were able to select the optimal DNA sequences bound by the individual POU homeodomain and POU-specific domain. These sites generally correspond to the two halves of the Oct-1 recognition element, the octamer sequence (ATGCA and AAAT). Together, these experiments demonstrated that the POU domain is comprised of two structurally independent domains, each of which is responsible for recognition of one half of the octamer sequence.

The x-ray crystal structure of the Oct-1 POU domain—octamer (ATGCAAAT) complex (which will be described in detail in Chapter 2) confirms the biochemical evidence that the POU domain is made up of two structurally autonomous subdomains (Figure 1). The POU-specific domain and the POU homeodomain are each docked in the major groove of the double helix, but they are on opposite sites of the DNA and there are no protein-protein contacts between them in the complex. The structure of the POU-specific domain is similar to the DNA-binding domains of λ and 434 repressor and 434 cro (first shown by Assa-Munt et al., 1993 and Dekker et al., 1993), while the structure of the POU homeodomain is nearly identical to the known structures of other homeodomains. Each subdomain makes base contacts with a 4 - 5 base pair subsite: The POU-specific domain contacts an ATGCA subsite in the 5' portion of the octamer sequence and the POU homeodomain contacts an AAAT subsite in the 3' portion of this sequence. The 24 residue linker connecting the subdomains is not visible in the electron density maps, indicating it is very flexible. Prompted by features of the crystal structure, subsequent biochemical experiments have revealed that the two Oct-1 POU subdomains bind DNA cooperatively, in the absence of the linker. These studies have also allowed quantitation of the effective concentration

provided by the linker. [Cooperativity in Oct-1 POU domain—DNA interactions will be discussed in Chapter 3.] Thus, the POU domain is comprised of structurally independent subdomains whose binding is thermodynamically coupled.

TFIIIA-type zinc fingers and paired domains also contain linked DNA-binding modules

Studies of Oct-1 POU domain—DNA interactions have provided both a structural and energetic description of how its POU-specific domain and POU homeodomain cooperate to recognize the octamer site. Like the POU domain, crystal structures of TFIIIA-type zinc finger—DNA complexes and a paired domain—DNA complex reveal that they also comprised of structurally independent modules. How are covalently linked modules used in DNA recognition by these motifs? Specifically, the following questions can be raised:

- How do the individual subdomains contribute to DNA-binding affinity and specificity?
- What is the contribution of the linker connecting the subdomains? Does it influence the positioning the domains on the DNA? Or is it simply a covalent tether?
- Do the subdomains interact when bound to DNA?

The following discussions of zinc fingers and paired domains summarize the biochemical and structural data addressing these issues. A comparison of these three motifs - POU domains, zinc fingers, and paired domains - allows us to understand the different ways in which linked modules can be used in DNA recognition.

Zinc fingers

Zinc finger domains (of the type first identified in transcription factor IIIA (TFIIIA)) contain tandem repeats of a conserved 30-residue zinc finger sequence. The number of fingers in these proteins can vary considerably - from a few as two to as many as 37 repeats have been seen (it is not clear, however, how many of these repeats bind DNA). NMR studies first revealed that an individual finger module is comprised of a two-stranded β-sheet and a single α-helix, which are held together by a zinc ion tetrahedrally coordinated by the invariant cysteine and histidine residues (Párraga et al., 1988; Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990). The high degree of sequence homology among these repeats strongly suggests that all TFIIIA-type zinc finger modules contain this structural framework.

Adjacent zinc fingers are connected by a linker of 4-5 residues, which in many proteins has the consensus TGEK. To understand the roles of these conserved linker residues, two groups (Choo and Klug 1993; Clemens et al. 1994) have made mutations in the linker connecting fingers 1 and 2 of TFIIIA, which contains the TGEK consensus. Point mutations at any position decrease binding at least 10-fold, and some decrease binding >300-fold. Insertion of a single serine residue between the G and E residues or between the E and K residues reduces binding about 10-fold (Choo and Klug, 1993). These data suggest that the linker actively contributes to DNA recognition by zinc fingers. However, it is important to note that some zinc finger proteins contain rather different linkers. For example, the linker connecting the two fingers in the Tramtrack protein contains the divergent sequence KRNVK. It is possible that the role of the linker may vary in different zinc finger

proteins.

Many biochemical studies have been directed toward assessing the contribution of individual zinc finger modules to DNA binding. Frankel et al. (1987) showed that a peptide corresponding to a single finger from TFIIIA can fold independently, but cannot bind DNA in a sequence-specific manner. However, the two zinc fingers from the Tramtrack protein can bind DNA specifically, albeit with rather low affinity ($K_d \sim 10^{-7}$ M) (Fairall et al., 1992). These results suggest that a minimum of two fingers is required for DNA binding. However, in proteins with a large number of fingers, it appears that they do not all contribute equally to DNA binding. One system that has been studied extensively is the DNA-binding domain of TFIIIA, which contains nine zinc fingers and binds the 45 base pair site in the internal control region (ICR) of the 5S RNA gene. Liao et al. (1992) and Clemens et al. (1994) have measured binding constants for a series of truncated TFIIIA zinc finger proteins and their studies indicate that the first three fingers account for 95% of the binding energy of the entire nine-finger protein. Thus, the binding contribution resulting from an increase in the number of fingers is not simply additive. Additional studies of this system suggest that fingers 1 -3 and 7- 9 may make all of the major groove DNA contacts, while fingers 4 - 6 may simply serve as a spacer element between these two sets of fingers (Vrana et al., 1988; Hayes and Clemens, 1992; Hayes and Tullis, 1992); therefore, some fingers may play indirect roles in DNA recognition.

X-ray crystal structures of three different zinc finger—DNA complexes have been solved, allowing a detailed analysis of the arrangement of the modules on DNA and of the contacts they make with their recognition elements. The Tramtrack complex contains two fingers (Fairall et al., 1993), the Zif268 complex contains three fingers (Figure 2) (Pavletich and Pabo,

1991), and the GLI complex contains five fingers (Pavletich and Pabo, 1993). These structures reveal that neighboring fingers are arranged in a way that reflects the helical pitch of the DNA: a rotation of approximately 96° around the DNA axis and a translation of approximately 10 Å superimposes one finger with the next (Pavletich and Pabo, 1991). Individual fingers generally make base contacts to three or four base pair subsites. The linkers residues are well ordered in the Zif268 and GLI complexes, suggesting that they could play a role in positioning the fingers on the DNA. In contrast, the residues of the Tramtrack linker have high thermal factors, and have different conformations in the two complexes of the asymmetric unit. The only DNA contact made by a linker residue in any of the three structures is a phosphate contact made by a lysine in one of the GLI linkers. Adjacent fingers in all three complexes are structurally independent - at most, a single protein-protein contact is made between neighboring modules on the DNA. A striking exception is the GLI—DNA complex, where finger 1 does not contact the DNA, but instead makes extensive protein-protein interactions with the DNA-distal side of finger 2. This observation supports the biochemical evidence that some of the fingers in zinc finger domains may play auxiliary roles in DNA recognition.

The paired domain

Recent structural studies confirm the biochemical evidence that the paired domain, like the POU domain, is a bipartite motif. The 128 amino acid paired box was first identified as a region of homology in the Drosophila *paired* (*prd*) and *gooseberry* genes (Bopp et al., 1986) and was subsequently shown to encode a sequence-specific DNA-binding domain (Treisman et al., 1991;

Chalepakis et al., 1991). [Many proteins containing a paired domain also contain a homeodomain, and these two domains can bind DNA simultaneously; however, the physiological relevance of this observation is unknown. (Treisman et al., 1991).] Initial characterization of the *prd* paired domain showed that the first 90 amino acids are sufficient for DNA binding. (Treisman et al., 1991). This is consistent with the observation that all known developmental missense mutations in the paired box of Pax genes (the mammalian paired box-containing genes) map to the N-terminal subdomain, which is more highly conserved among paired family members that the C-terminal domain (Walther et al., 1991). Although the C-terminal half of the paired domain is apparently dispensable for *prd* binding, biochemical and genetic evidence strongly suggest that the C-terminal region is required for optimal DNA binding by other paired domain family members.

To investigate the role of the C-terminal residues of another paired domain, Czerny et al. (1993) have compared binding by the intact Pax5 protein and binding by a deletion variant of the protein lacking the C-terminal 36 residues of the paired domain. The deletion variant bound only a subset of the naturally occurring Pax5 sites tested. Comparison of the sites the variant bound with those it did not bind allowed a consensus site to be deduced which consisted of two half sites. Those sites containing a perfect consensus in the 3' half of the site were recognized by the protein lacking the C-terminal residues of the paired domain. However, the C-terminal region is required for binding to DNA sites which lack a 3' consensus. These experiments were the first to suggest the bipartite nature of the paired domain and showed that, for Pax5, different DNA sites have different requirements for the two halves of the domain.

Detailed analysis of the DNA interactions of the Pax6 protein and of

a splicing variant of the protein, Pax6-5a, further supports the observation that the role N-terminal and C-terminal halves of the paired domain can depend on the nature of the DNA site. The optimal DNA site selected for wild-type Pax6 protein ("P6CON") bears no obvious similarity to the optimal DNA binding site selected for Pax6-5a ("5aCON"), which has an extra 14 amino acids inserted into the N-terminal region (Epstein et al., 1994). The 5aCON site consists of two imperfect repeats and Pax6-5a binds this site as a dimer, while Pax6 binds as a monomer. Interestingly, Pax6-5a can bind to DNA containing a single 5aCON half site while Pax6 cannot, suggesting that there may be significant differences in how the C-terminal domains of these proteins are docked on the DNA. Deletion analysis of these proteins demonstrated that the C-terminal region is critical for binding of either form of the protein to 5aCON DNA site while an intact N-terminal region is necessary and sufficient for binding to the P6CON DNA site; thus, Pax6-5a does not bind this site. The authors describe the insertion as a "molecular toggle to unmask the DNA-binding potential of the carboxyl terminus" (Epstein et al., 1994). These observations suggest that both the N-terminal and C-terminal halves of the Pax6 paired domain are capable of high affinity, sequence-specific DNA binding.

The crystal structure of the paired domain from the Drosophila *prd* protein complexed with DNA reveals that this "domain" is actually comprised of two structurally independent helical subdomains, which make no protein-protein contacts with one another (Figure 3) (Xu et al., 1995). In the crystallized complex, the 60 residue N-terminal subdomain (which also contains a 2-stranded β sheet at its N-terminus) makes extensive contacts with the DNA while the 40 residue C-terminal helical subdomain does not contact the DNA at all. This is consistent with the biochemical evidence that

the C-terminal half of the paired domain is not required for DNA recognition by the *prd* protein (Treisman et al., 1991). Additionally, binding site selections confirm that the DNA site used in the crystallized complex is sufficient for high affinity, sequence-specific binding by the protein (Xu et al., 1995). The ~20 amino acid linker connecting the domains lies in the minor groove of the DNA and main chain atoms make base and phosphate contacts. Structural studies of a Pax6—DNA complex are currently underway (a collaboration between the Pabo lab and Richard Maas' lab) and are likely to reveal how the C-terminal paired subdomain interacts with DNA.

Comparison of covalently-linked DNA-binding domains

The data for these three systems - POU domains, zinc fingers, and paired domains - reveals that although there are some similarities, there are important differences in how these motifs use linked DNA-binding modules in recognition. Perhaps the key similarity among these three families is that their subdomains are structurally independent and make few or no contacts with one another when complexed with DNA: there is at most one contact between adjacent zinc fingers, and the POU and paired subdomains do not contact one another at all. However, it is clear that these motifs differ significantly in other respects. In particular, it is interesting to compare: 1) the contribution of individual subdomains to DNA binding and 2) the role of the covalent linker connecting the modules.

The data described above allow us to compare how individual subdomains contribute to DNA binding among these three families. For the POU domain, both the POU-specific domain and the POU homeodomain make many DNA contacts and both are required for optimal DNA binding by

all POU proteins that have been tested. Experiments with peptides corresponding to the individual subdomains reveal that the POU homeodomain binds DNA with relatively low affinity and specificity and POU-specific domain binding is barely detectable, but in the context of the intact POU domain, they cooperate to achieve high affinity DNA binding. For the paired domain, the contributions of the N- and C-terminal subdomains appear to depend on the particular protein and on the nature of the DNA site. Thus, it has been shown for different paired domain proteins on different sites that: 1) the N-terminal subdomain can be sufficient for binding (e.g., *prd*); 2) the C-terminal subdomain can be sufficient for binding (e.g., Pax6-5a); and 3) both the N- and C-terminal subdomains can be required for binding (e.g., Pax5). A careful thermodynamic analysis of the contributions of the individual paired subdomains to DNA binding would provide a great deal of insight into DNA recognition by this motif. Finally, evaluating the contribution of individual zinc fingers to DNA binding is somewhat more complicated because different proteins contain different numbers of fingers, and it appears that there is not a simple correspondence between the number of fingers in a protein and its DNA-binding affinity. For example, the three fingers of Zif268 have about the same affinity for their binding site (K_d ~6 nM) (Pavletich and Pabo, 1991) as the five fingers of GLI have for their binding site (K_d ~20 nM) (Pavletich and Pabo, 1993). Examining the structures of these two complexes reveals that, while all three fingers of Zif268 make multiple base and phosphate contacts, fingers 4 and 5 of the GLI peptide make all but one of the specific base contacts. These observations provide a structural explanation for the nonequivalent contributions zinc fingers can make to DNA binding.

The linkers connecting the DNA-binding subdomains in these

three families differ significantly in their length, flexibility, and in their contribution to DNA binding. The linker connecting the POU-specific domain and the POU homeodomain is quite variable among POU family members: it ranges in length from 15 residues (Pit-1) to 56 residues (CEH-18) and the amino acid sequences vary dramatically. Both biochemical and structural evidence suggests that this region is unstructured and may simply serve as a flexible tether to covalently link the subdomains. In contrast, the linkers connecting adjacent zinc fingers are short (4 - 5 residues) and their amino acid sequence is generally conserved, suggesting they may be involved in correctly positioning the fingers on the DNA. Like the zinc finger linkers, the linker connecting the N-terminal and C-terminal helical subdomains of the paired domain also has a defined length (~20 residues) and sequence among the paired family members. The *prd*—DNA crystal structure shows that the linker lies in the minor groove of the DNA and main chain atoms make base and phosphate contacts (Xu et al., 1995). Additional biochemical analysis will be necessary to understand the energetic contributions of these contacts and the roles of conserved residues.

DNA recognition with covalently linked DNA-binding modules

What are the advantages to using covalently linked modules in DNA recognition? Fundamentally, DNA-binding domains containing multiple subdomains recognize relatively long binding sites, so that a single protein molecule binds DNA with high affinity and specificity. For example, a homeodomain generally contacts a six base pair sequence while the POU-homeodomain plus the POU-specific domain contact an approximately 10 base pair sequence. Some zinc finger proteins can recognize sites at least 15

base pairs in length and the paired domain can contact sites on the order of 20 base pairs. As these proteins make a large number of specific contacts, they bind DNA with high affinity and readily distinguish their specific sites from nonspecific sites (see Ptashne, 1992, for a discussion of DNA-binding specificity).

Covalent linkage of DNA-binding modules ensures that their binding is highly cooperative, and this makes an important contribution to DNA binding by these proteins. When one domain is bound to DNA, the other domain(s) will be held near the DNA at a high local concentration, thus driving DNA binding. This is referred to as a chelate effect and can be described in terms of an effective concentration. It has been possible to quantitate the effective concentration between the Oct-1 POU-specific domain and the POU homeodomain. This value, 3.6 mM, is well above the dissociation constant of either individual domain; therefore, an intermediate complex with just one subdomain bound will only be present transiently. [The chelate effect will be discussed in further detail in Chapter 3.] In addition, our studies of the Oct-1 POU domain have revealed that binding by the POU-specific domain and the POU homeodomain is also coupled by overlapping DNA contacts made by the adjacently-bound subdomains. Features of the TFIIIA-type zinc finger—DNA complexes strongly suggest that binding by adjacent fingers may be coupled in a similar way.

Of course, many transcription factors bind DNA as dimers and higher order multimers, which also allows for a large number of contacts to be made to the DNA and also involves cooperative interactions between monomers. While cooperative binding by linked domains is mediated through covalent linkage, cooperative binding by noncovalent dimers is mediated through protein-protein interfaces. The evolution of transcription

factors containing multiple DNA-binding domains has resulted in proteins which can bind efficiently as monomers. One can imagine that the different zinc finger proteins were created by simple duplication of the small zinc-binding framework. For POU domains, a second DNA-binding domain has been linked to a homeodomain to create a composite DNA-binding motif. Finally, while the subdomains of the paired domain do not correspond precisely to domains used in other contexts, they both have similarity to the homeodomain and Hin recombinase (Xu et al., 1995).

With regard to discussion of the evolutionary origin of these modular DNA-binding domains, it is important to note that the genomic structures within these families are variable and there is not always a simple correspondence of one DNA-binding domain or subdomain per exon in all cases. Thus, in those POU domain proteins for which the genomic structure is known, the POU domain has been shown to be coded on 1, 2, or 4 exons (Messier et al., 1993). [In fact, some POU domain proteins have no introns (Verrijzer and van der Vliet, 1993)]. For Pax proteins, Pax3 and Pax7 contain an intron-exon boundary in the paired domain linker, while Pax2, 5, 6 and 8 have boundaries within N-terminal subdomain of the paired domain (Walther et al., 1991). For zinc fingers, there are proteins in which all fingers are coded on separate exons (such as the Wilms' tumor gene product (Haber et al., 1990)), but there are other proteins that have multiple fingers on a single exon and still others have an intron/exon boundary within a finger. Therefore, examination of the intron/exon structure does not provide reliable clues to the evolutionary origin of these domains.

Implications for the design of novel DNA-binding proteins

DNA-binding domains of linked modules offer distinct advantages for the design of DNA-binding proteins with altered specificities. Structural independence of the modules allows them to be varied individually and recombined in novel arrangements. Zinc fingers have been especially attractive for design purposes because they are able to recognize a diverse set of DNA sequences. Taking advantage of these observations, several groups have selected or designed zinc finger proteins with altered specificities. Rebar and Pabo (1994) have expressed the three fingers of Zif268 on the surface of filamentous phage and by randomizing the DNA-contacting residues in finger 1, they selected phage expressing peptides that preferentially bind DNA with a unique finger 1 subsite. Other groups have subsequently carried out similar experiments using phage display of zinc finger peptides (Jamieson et al., 1994; Choo and Klug, 1994a; Wu et al., 1995). Choo and Klug (1994b) have carried this strategy one step further and have recombined individually selected fingers to create a novel protein which binds a unique nine base pair DNA site. Strategies are currently being developed in the Pabo lab to optimize the design of multi-finger peptides which recognize entirely new DNA sites. Such strategies should prove key for the design of novel DNA-binding proteins for use in biotechnology and gene therapy.

It is interesting to note that there are naturally occurring examples of zinc finger "shuffling" to create proteins with new binding specificities, through alternative mRNA splicing. A well-characterized example of a zinc finger protein that is subject to developmentally-regulated splicing is the *Drosophila* transcription factor CF2. Alternatively-spliced transcripts of this protein encode three protein isoforms that differ in the number of zinc fingers (Hsu et al., 1992). DNA binding studies with two of the three isoforms reveal that they have distinct DNA-binding specificities (Gogos et al., 1992).

Other examples of zinc finger proteins with isoforms containing different numbers of fingers include the Evi-1 gene (Bordereaux et al., 1990) and the *C. elegans* tra-1 gene (Zarkower and Hodgkin, 1992). Although the physiological roles of the different isoforms of these zinc finger proteins is not yet understood, the modular nature of zinc finger domains clearly provides a mechanism for regulating the DNA-binding specificity of transcription factors by alternative splicing.

The characterization of modular DNA-binding domains has also prompted the design of a novel hybrid DNA-binding protein. Pomerantz et al. (1995) have engineered a molecule in which fingers 1 and 2 from Zif268 are fused to the Oct-1 POU homeodomain. Computer modeling showed that these modules could be aligned on DNA in such a way that the C-terminus of zinc finger 2 is 8.8 Å away from the N-terminus of the homeodomain (Figure 4). This suggested that a 4 amino acid linker could be used to connect the domains. The optimal DNA site for the chimeric protein was selected from a randomized pool of oligonucleotides and contained half sites which strongly resembled the optimal binding sites for the individual domains. The chimeric molecule strongly prefers this binding site over the binding sites of the parent molecules, and has a dissociation constant of 8.4×10^{-10} M. Fusion of a transcriptional activation domain to this chimeric protein allowed it to promote transcription in cells containing a reporter construct with the selected binding site. Thus, DNA-binding modules from different families can be recombined to create new hybrid molecules with unique specificities.

Conclusions

By designing proteins with covalently linked DNA-binding modules, nature has

created monomeric proteins which bind DNA with high affinity and specificity. The large number of DNA contacts made by these proteins allows efficient discrimination of a specific DNA site from nonspecific sites, and the covalent linkage ensures that the binding of modules is inherently cooperative. The three families of modular DNA-binding domains discussed in this chapter - POU domains, TFIIIA-type zinc fingers, and paired domains - have rather different properties with regard to the contribution of individual subdomains to DNA-binding and the role of the linker connecting the subdomains. These differences are intriguing and continued studies of these motifs will improve our understanding of these issues.

In the following chapters, the structural and biochemical studies of the POU domain from the transcription factor Oct-1 will be described. These studies have provided detailed insight into DNA interactions by this bipartite motif. The x-ray crystal structure (described in Chapter 2) has revealed how the POU subdomains - the POU-specific domain and the POU homeodomain - cooperate to recognize a specific DNA sequence. Biochemical studies (described in Chapter 3) have measured the energetic contribution of each subdomain and the linker polypeptide to DNA binding. Additionally, these studies have allowed measurement of the coupling energy between the subdomains. Chapter 4 suggests directions for future research regarding the POU domain.

References

Ades, S. E. (1995). Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Assa-Munt, N., Mortishire-Smith, R. J., Aurora, R., Herr, W., and Wright, P. E. (1993). The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. Cell *73*, 193-205.

Aurora, R., and Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. Mol. Cell. Biol. *12*, 455-467.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of Drosophila. Cell *47*, 1033-1040.

Bordereaux, D., Fichelson, S., Tambourin, and Gisselbrecht, S. (1990). Alternative splicing of the Evi-1 zinc finger gene generates mRNAs which differ by the number of zinc finger motifs. Oncogene *5*, 925-927.

Botfield, M. C., Jancso, A., and Weiss, M. A. (1992). Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. Biochemistry *31*, 5841-5848.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991). The molecular basis of the *undulated/Pax-1* mutation. Cell *66*, 873-884.

Choo, Y. and Klug, A. (1993). A role in DNA binding for the linker sequences of the first three fingers of TFIIIA. Nucl. Acids Res. *21*, 3341-3346.

Choo, Y., and Klug, A. (1994a). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. USA *91*, 11163-11167.

Choo, Y., Sánchez-García, I., and Klug, A. (1994b). *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature *372*, 642-645.

Clemens, K. R., Zhang, P., Liao, X., McBryant, S. J., Wright, P. E., and Gottesfeld, J. L. (1994). Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. J. Mol. Biol. *244*, 23-35.

Czerny, T., Schaffner, G., and Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. Genes Dev. *7*, 2048-2061.

Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., van der Vliet, P. C., and Kaptein, R. (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. Nature *362*, 852-855.

Epstein, J. A., Glaser, T., Cai, J., Jepeal, L, Walton, D. S., and Maas, R. L. (1994). Two independent and interactive DNA-binding subdomains of the Pax-6 paired domain are regulated by alternative splicing. Genes Dev. *8*, 2022-2034.

Fairall, L., Harrison, S. D., Travers, A. A., and Rhodes, D. (1992). Sequence-specific DNA binding from the *Drosophila melanogaster* Tramtrack protein. J. Mol. Biol. *226*, 349-366.

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature *366*, 483-487.

Finney, M., Ruvkin, G., and Horvitz, H. R. (1988). The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein containing a homeo domain and an extended sequence similarity to mammalian transcription factors. Cell *55*, 757-769.

Frankel, A. D., Berg, J. B., and Pabo, C. O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. USA 84, 4841-4845.

Frankel, A. D. and Kim, P. S. (1991). Modular structure of transcription factors: implications for gene regulation. Cell *65*, 717-719.

Gogos, J. A., Hsu, T., Bolton, J., and Kafatos, F. C. (1992). Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. Science *257*, 1951-1955.

Haber, D. A., Buckler, A. J., Glaser, T., Call, K. M., Pelletier, J., Sohn, R. L., Douglass, E. C., and Housman, D. E. (1990). An internal deletion within a 11p13 zinc finger gene contributes to the development of Wilms' tumor. Cell *61*, 1257-1269.

Hayes, J. J., and Clemens, K. L. (1992). Locations of contacts between individual zinc fingers of *Xenopus laevis* transcription factor IIIA and the internal control region of a 5S RNA gene. Biochemistry *31*, 11600-11605.

Hayes, J. J., and Tullius, T. D. (1992). Structure of the TFIIIA-5 S DNA complex. J. Mol. Biol. *227*, 407-417.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A.,

Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The POU domain: a large conserved region in the mammalian *pit-1, oct-1, oct-2,* and *Caenorhabditis elegans unc-86* gene products. Genes Dev. *2,* 1513-1516.

Hsu, T., Gogos, J. A., Kirsh, S. A. and Kafatos, F. C. (1992). Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene. Science *257,* 1946-1950.

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1—Pit-1 interactions. Cell *61,* 1021-1033.

Jamieson, A. C., Kim, S. -H., and Wells, J. A. (1994). In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry *33,* 5689-5695.

Johnson, W. A., and Hirsh, J. (1990). Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. Nature *343,* 467-470.

Klevit, R. E., Herriott, J. R., and Horvath, S. J. (1990). Solution structure of a zinc finger domain of yeast ADR1. Proteins *7,* 215-226.

Kristie, T. M., and Sharp, P. A. (1990). Interactions of the Oct-1 POU subdomains with a specific DNA sequence and with the HSV α-*trans*-activator protein. Genes & Dev. *4,* 2383-2396.

Lee, M. S., Gippert, G. P., Sonman, K. V., Case, D. A., and Wright, P. E. (1989). Three dimensional solution structure of a single zinc finger DNA-binding domain. Science *245,* 635-637.

Liao, X., Clemens, K. R., Linda, T., Wright, P. E., and Gottesfeld, J. M. (1992). Specific interaction of the first three zinc fingers of TFIIIA with the internal control region of the *Xenopus* 5S RNA gene. J. Mol. Biol. *223,* 857-871.

Messier, H., Brickner, H., Geiwad, J., and Fotedar, A. (1993). A novel POU domain protein which binds to the T-cell receptor β enhancer. Mol. Cell. Biol. *13,* 5450-5460.

Ominchinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K., and Gronenborn, A. M. (1990). High-resolution three-dimensional structure of a single zinc-finger from a human enhancer binding protein in solution. Biochemistry *29,* 9324-9334.

Pabo, C. O., and Sauer, R. T. (1992). Transcription factors: structural families and

principles of DNA recognition. Annu. Rev. Biochem. *61*, 1053-1095.

Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., and Klevit, R. E. (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. Science *241*, 1489-1492.

Pavletich, N. P. and Pabo, C. O. (1991). Zinc finger—DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science *252*, 809-817.

Pavletich, N. P., and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701-1707.

Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995). Structure-based design of transcription factors. Science *267*, 93-96.

Ptashne, M. (1992). A Genetic Switch. Cell Press, Cambridge, MA.

Rebar, E. J. and Pabo, C. O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science *263*, 671-673.

Rosenfeld, M. G. (1991). POU-domain transcription factors: pou-er-ful developmental regulators. Genes & Dev. *5*: 897-907.

Ruvkin, G., and Finney, M. (1991). Regulation of transcription and cell identity by POU domain proteins. Cell *64*, 475-478.

Sturm, R. A., and Herr, W. (1988). The POU domain is a bipartite DNA-binding structure. Nature *336*, 601-604.

Tjian, R., and Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell *77*, 5-8.

Treisman, J., Harris, E., and Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the Paired homeo domain proteins. Genes Dev. *5*, 594-604.

Verrijzer, C. P., Kal, A. J., and van der Vliet, P. C. (1990). The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. Genes Dev. *4*, 1964-1974.

Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., van Leeuwen, H. C., Strating, M. J. J., and van der Vliet, P. C. (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. EMBO J. *10*, 4993-5003.

Verrijzer, C. P., and van der Vliet, P. C. (1993). POU domain transcription factors. Biochim Biophys. Acta. *1173*, 1-21.

Vrana, K. E., Churchill, M. E. A., Tullius, T. D., and Brown, D. D. (1988). Mapping functional regions of transcription factor TFIIIA. Mol. Cell Biol. *8*, 1684-1696.

Walther, C., Guenet, J -L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. (1991). Pax: a murine multigene family of paired box-containing genes. Genomics *11*, 424-434.

Wu, H. Yang, W -P., and Barbas, C. F. (1995). Building zinc fingers by selection: toward a therapeutic application. Proc. Natl. Acad. Sci. USA. *92*, 344-348.

Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. Cell *80*, 639-650.

Zarkower, D., and Hodgkin, J. (1992). Molecular analysis of the C. elegans sex-determining gene tra-1: a gene encoding two zinc finger proteins. Cell *70*, 237-249.

Figure 1. Overview of the Oct-1 POU domain—DNA complex. The POU homeodomain is colored red, the POU-specific domain is colored yellow, and the DNA double helix is colored blue. A linker (pink) has been included to emphasize that the two domains are linked; however, these 24 residues cannot be seen in the electron density maps.

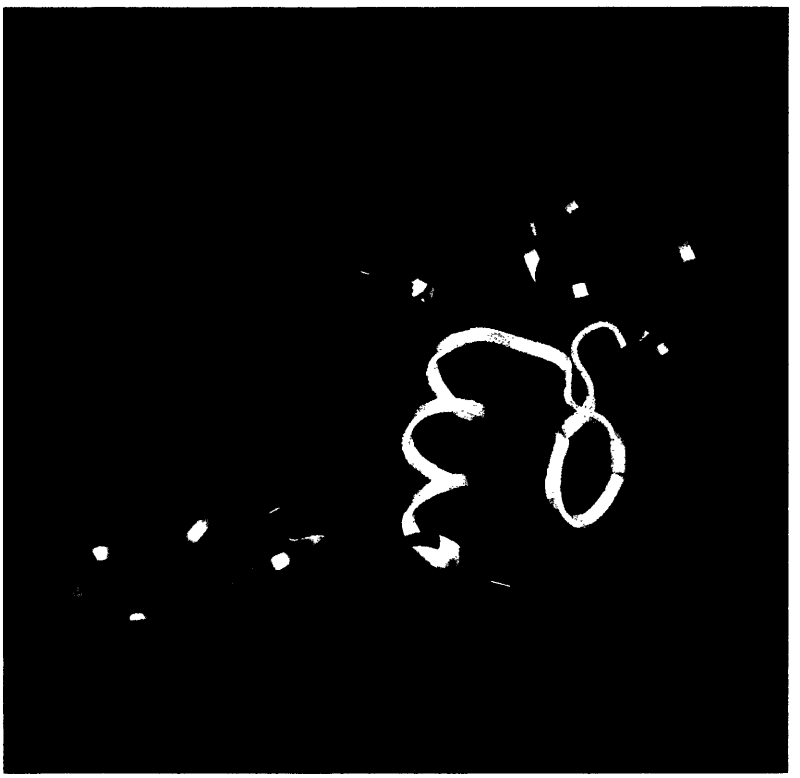

Figure 2. Overview of the zif268—DNA complex (Pavletich and Pabo, 1991). The three fingers are colored red, yellow, and green, and the linkers connecting the fingers are pink. The DNA double-helix is colored blue.

Figure 3. Overview of the *prd* paired domain—DNA complex (Xu et al., 1995). The N-terminal subdomain is colored red, the C-terminal subdomain is colored yellow, and the linker connecting them is colored pink. The DNA double helix is colored blue.

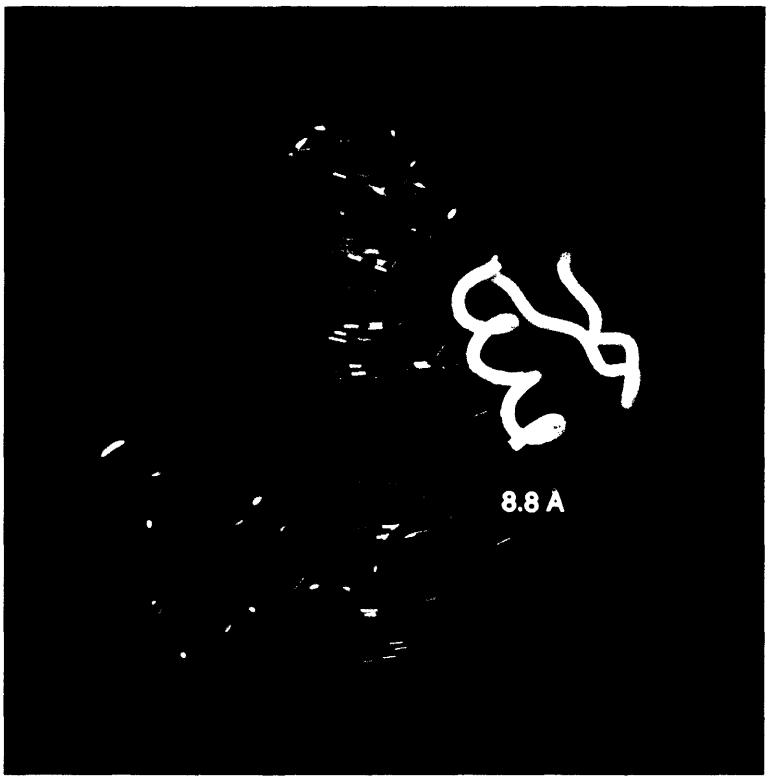

Figure 4. Model of a zinc finger—homeodomain hybrid (figure taken from Pomerantz et al., 1995). The two fingers from zif268 are colored pink and yellow, and the Oct-1 POU homeodomain is colored red. The DNA is blue, with the base pairs of each domain's subsite highlighted in cyan. The distance between the C-terminus of finger 2 and the N-terminus of the POU homeodomain is 8.8 Å.

CHAPTER 2

Crystal Structure of an Oct-1 POU Domain—Octamer Complex: DNA Recognition with Tethered DNA-Binding Modules

Summary

The structure of an Oct-1 POU domain—octamer DNA complex has been solved at 3.0 Å resolution. The POU-specific domain contacts the 5' half of this site (ATGCAAAT), and - as predicted from NMR studies - the structure, docking, and contacts are remarkably similar to the λ and 434 repressors. The POU-homeodomain contacts the 3' half of this site (ATGCAAAT), and the docking is similar to that of the engrailed, MATα2, and Antennapedia homeodomains. The linker region is not visible and there are no protein-protein contacts between the domains, but overlapping phosphate contacts near the center of the octamer site may favor cooperative binding. This novel arrangement raises important new questions about cooperativity in protein-DNA recognition.

Introduction

The POU domain, a DNA-binding domain which characterizes a family of eukaryotic transcription factors, has a novel modular structure that raises important questions about protein-DNA recognition and gene regulation. This domain contains a 75 amino acid POU-specific domain, a variable linker of 15 to 30 residues, and a 60 amino acid POU-type homeodomain (Herr et al., 1988). Previous studies have shown that both the POU-specific domain and the POU homeodomain can fold as independently stable structures (Botfield et al., 1992), that both domains contact the DNA in the complex (Verrijzer et al., 1992; Aurora & Herr, 1992), and that both domains are required for high-affinity, sequence-specific binding (Sturm & Herr, 1988; Ingraham et al., 1990; Verrijzer et al., 1990). Recent NMR studies (Assa-Munt et al., 1993; Dekker et al., 1993) have shown that the structure of the N-terminal POU-specific domain is closely related to the DNA-binding domains of the phage λ and 434 repressors and 434 cro protein (Jordan & Pabo, 1988; Beamer & Pabo, 1992; Aggarwal et al., 1988; Wolberger et al., 1988; Mondragón & Harrison, 1991). Sequence comparisons and structural studies of other homeodomains had provided a plausible model for the folding and docking of the POU homeodomain (Qian et al., 1989; Otting et al., 1990; Kissinger et al., 1990; Wolberger et al., 1991). Without structural studies of the POU domain—DNA complex, however, it had not been clear how reliable these modeling efforts are or how the individual domains interact with each other and cooperate in DNA binding.

Over 20 members of the POU family have been identified, and these proteins are involved in various aspects of transcriptional regulation (reviewed in Ruvkun & Finney, 1991; Rosenfeld, 1991). For example, Pit-1 [also referred to as GHF-1 (Bodner et al., 1988)] is expressed in the pituitary and regulates

transcription of prolactin and growth hormone genes (Nelson et al., 1988; Ingraham et al., 1988; Li et al., 1990). In contrast, Oct-1 is broadly expressed in mammals and regulates transcription of small nuclear RNA genes and the histone H2B gene by binding to the octamer site ATGCAAAT (reviewed in Herr, 1992). Oct-1 also acts at TAATGARAT sites (R=purine), forming a multi-protein complex with the herpes simplex virus protein VP16 and another cellular factor (HCF) that activates the viral immediate-early genes (Gerster & Roeder, 1988; Kristie et al., 1989; Stern et al., 1989). Many POU domain proteins function as monomers, but others - such as Pit-1 - commonly form dimers at their target DNA sites (Ingraham et al., 1990). Interestingly, Oct-1 binds as either a monomer or a dimer, depending on the sequence of the regulatory element (Kemler et al., 1989; Poellinger & Roeder, 1989).

To understand precisely how the POU-specific domain and the POU homeodomain bind to DNA and how they interact with each other in the complex, we crystallized the Oct-1 POU domain (Figure 1) bound to a DNA sequence that contains an octamer site and flanking base pairs from the human histone H2B promoter (Figure 2). The structure of this complex has been solved at 3.0 Å resolution, revealing both the overall arrangement of the complex and details of the extensive base and phosphate contacts that each domain makes with the DNA. Here we describe the crystal structure of the POU domain—DNA complex, compare the individual domains with the phage repressors and with other homeodomains, and discuss the broader implications for understanding cooperativity in protein-DNA recognition.

Results and Discussion

Overall Arrangement of the POU Domain—DNA Complex

In our crystal structure, both the POU-specific domain and the POU homeodomain bind directly to the DNA and make extensive base and phosphate contacts (Figures 3 and 9). The POU-specific domain makes its primary contacts with the ATGC sequence in the 5' half of the octamer site, and the POU homeodomain makes its primary contacts with the AAAT sequence in the 3' half of the octamer site. The crystal structure of the complex shows that these domains bind on opposite sides of the double helix (Figures 3 and 4). There are no protein-protein contacts between the POU-specific domain and the POU homeodomain but these domains do make some contacts to overlapping phosphates (Figure 9). The 24 residue linker that connects the POU-specific domain and the POU homeodomain cannot be seen in the electron density maps, indicating that it is disordered in the crystal. In binding DNA, the individual domains each use a helix-turn-helix (HTH) unit for recognition and helix 3 from each domain is positioned in the major groove of the DNA. The POU-specific domain has a pair of conserved glutamines in the HTH unit (at the start of helix 2 and at the start of helix 3), and these make contacts that are exactly analogous to contacts seen in the λ and 434 complexes (Pabo et al., 1990). The POU homeodomain is very similar to previously characterized homeodomains, both in structure and DNA binding. This structure provides the first opportunity to see the N-terminal arm of a homeodomain in the context of a larger polypeptide and our structure confirms that this N-terminal arm binds in the minor groove.

Structure and DNA Contacts of the POU-Specific Domain

Our structure reveals that the docking of the POU-specific domain and even some details of the side chain-base interactions are closely related to features seen in the λ and 434 repressor—DNA complexes. As shown from the NMR studies, the POU-specific domain contains four α-helices, with helices 2 and 3

forming a variant HTH unit. Helix 2 and the turn are each three residues longer than in the phage repressors (Assa-Munt et al., 1993; Dekker et al., 1993). Helix 3 is positioned in the major groove and makes all of the base contacts with the ATGC subsite (Figures 3, 4, and 5). Gln-44, which is the first residue of helix 3, makes two hydrogen bonds to the adenine at base pair 1, accepting a hydrogen bond from the N6 and donating a hydrogen bond to the N7 (Figure 6). This interaction is stabilized by Gln-27 (the first residue in helix 2), which hydrogen bonds to the side chain -NH of Gln-44 and also hydrogen bonds to a phosphodiester oxygen of adenine 1. This hydrogen bonding network is further stabilized by interactions with two other residues. Glu-51, (near the C-terminal end of helix 3) hydrogen bonds to Gln-27, while Arg-20 (from helix 1) hydrogen bonds to Glu-51 and also contacts a phosphodiester oxygen of thymine at base pair -1. These four residues (Arg-20, Gln-27, Gln-44, and Glu-51) are conserved in all of the known POU-specific domain sequences (Assa-Munt et al., 1993) and identical residues are found at corresponding positions of the 434 repressor and 434 cro—DNA complexes.

Thr-45 and Arg-49, which are also in helix 3, make additional base contacts with the ATGC subsite (Figure 5). Thr-45 donates a hydrogen bond to the O4 of thymine 2 and simultaneously accepts a hydrogen bond from the N4 of cytosine 3'. Serine is the only other residue found at this position in the POU domain (Assa-Munt et al., 1993), and we presume that it makes similar contacts. Arg-49, which is invariant in all known POU domains, hydrogen bonds to the O6 of guanine 3 and to the O6 and N7 of guanine 4'. These contacts must be important for binding because changing residue 49 to an alanine severely disrupts DNA binding by the Oct-1 POU domain (Cleary & Herr, 1995).

The POU-specific domain makes an extensive set of phosphate contacts, including some contacts in the AAAT subsite (Figure 9). Residues from

each of the four α-helices and residues from two of the loops are involved in making these backbone contacts. Arg-20, Thr-26, Gln-27 and Ser-48 contact phosphodiester oxygens along one strand of the DNA while Ser-43, Thr-46, Ser-56, Asn-59 and Lys-62 contact the other strand. Consistent with their expected importance in DNA binding, the residues that make phosphate contacts are all highly conserved or invariant among POU proteins.

It is interesting to note that the NMR studies of Dekker et al. (1993) gave very reliable information about the DNA contacts made by the Oct-1 POU-specific domain. Based on the loss of signal in the presence of DNA, Gln-27, Gln-44, Arg-49 and Asn-59 were correctly predicted to contact the DNA. Although the correlation (at other positions) between their predictions and the contacts seen in the crystal structure is not perfect, the overall success of this approach is quite encouraging.

NMR studies had shown that the structure of the POU-specific domain was similar to the structure of the λ and 434 repressors (Assa-Munt et al., 1993; Dekker et al., 1993). Noting this extensive structural similarity, which includes a pair of conserved glutamines, and the similarity of the Oct-1 and λ repressor binding sites allowed Assa-Munt et al. (1993) to correctly predict how the POU-specific domain would dock against the ATGC subsite. Indeed, comparing the crystal structure of the POU-specific complex with one half of the λ repressor complex shows that these proteins do bind DNA in a very similar manner (Figure 10) and that specific DNA contacts are conserved. Each protein has a glutamine at the start of helix 2 and a glutamine at the start of helix 3 and these make precisely analogous base and phosphate contacts in the POU complex and in the repressor complexes. Remarkably, even the extended hydrogen bonding network that involves Arg-20 and Glu-51 in Oct-1 is present in the 434 complexes. These similarities seem especially striking when we consider the vast

evolutionary separation between these proteins and when we note that Oct-1 has six additional residues in the middle of the HTH unit.

Structure and DNA Contacts of the POU Homeodomain

The overall structure and docking of the POU homeodomain are similar to those observed in other homeodomain—DNA complexes (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991). The three α-helices of the POU homeodomain form a compact globular structure, with helices 2 and 3 forming a HTH unit. The first nine residues of the homeodomain form an extended N-terminal arm that fits into the minor groove of the DNA.

Helix 3 of the POU homeodomain is docked in the major groove and contacts several bases in the AAAT subsite (Figures 3, 4, and 7). Critical contacts are made by the side chains of Val-47, Asn-51, and Cys-50. Val-47 is in van der Waals contact with the methyl group of thymine 8. Asn-51 is juxtaposed with adenine 7 and may hydrogen bond with the N6 and N7 positions. In the current structure, these hydrogen bonds would be 3.4 and 3.5 Å long, and clearly a higher resolution study will be needed to understand the details of these asparagine-adenine interactions and to determine whether any water molecules are present at this interface (see Qian et al., 1993). Cys-50 is of particular interest because this residue is characteristic of POU-type homeodomains. In the crystal, we find that the -SH group of Cys-50 makes a hydrogen bond to the main chain carbonyl oxygen of residue 46, which is in the preceding turn of the α-helix. Cys-50 also is in van der Waals contact with the methyl groups of thymine 9' and thymine 10', which are located just outside of the highly conserved octamer sequence. It is possible that Gln-54 also makes a base contact (perhaps to adenine 8) but the electron density is poorly defined for this residue, and it appears to be disordered in the complex.

As observed in other homeodomain—DNA complexes, the N-terminal arm of the homeodomain fits in the minor groove of the DNA, making contacts near the 5' end of the AAAT subsite. There is at least one base contact in the minor groove: Arg-5 is in position to donate a hydrogen bond to the N3 of adenine 5 (Figures 7 and 9). It is possible that Lys-3 also makes a contact in the minor groove, but we do not see clear electron density for the side chain.

Conserved residues of the POU homeodomain make a number of phosphate contacts (Figure 9). The side chains of Arg-5, Thr-6 and Arg-13 make hydrogen bonds to phosphodiester oxygens on the ATGCAAAT strand and the main chain -NH groups of Lys-3 and Thr-6 also contact phosphates of this strand. Phosphodiester oxygens on the other strand are contacted by the side chains of Lys-25, Ser-28, Arg-46 and Arg-53. Lys-55 and Lys-57, which make conserved phosphate contacts in the engrailed and MATα2 complexes, are positioned to allow phosphate contacts in the Oct-1 complex but these side chains do not have well-defined density in the current electron density maps.

The structure of the POU homeodomain—DNA complex confirms and extends the general model for homeodomain—DNA interactions. Although the amino acid sequence of the POU homeodomain is quite divergent - only 25% identity with the sequence of the engrailed homeodomain and 21% identity with the MATα2 homeodomain - we find that the Cα's for residues 9-58 of these homeodomains can be superimposed with an rms difference of less than 1 Å. Aligning the proteins in this way also gives a very good superposition of their DNA binding sites, showing that these homeodomains dock against the DNA in very similar ways (Figure 11). The POU homeodomain also makes several of the DNA contacts that are conserved in the engrailed and MATα2 complexes. The asparagine at position 51, which is conserved in all known homeodomains, interacts with the adenine at the third position of the AAAT subsite. Phosphate

contacts by Arg-53 and by residue 25 (which is a lysine in Oct-1 but a tyrosine in engrailed and MATα2) also are conserved. Many of the other DNA contacts made by the POU homeodomain have counterparts in either the engrailed complex or the MATα2 complex.

Residues of the Linker

The 24 residue linker that connects the POU-specific domain and the POU homeodomain cannot be seen in the electron density maps, suggesting that this region is disordered in the crystal and that it may be a flexible, unstructured part of the protein. This interpretation is consistent with a number of other observations: 1) Comparing POU domain proteins shows that the length and sequence of the linker region are highly variable. For example, Oct-1 and Oct-2 have the same DNA-binding specificity and have nearly identical POU-specific and POU homeodomains, but their linkers are very different (Herr et al., 1988). 2) Insertion of 6 alanine residues into the linker of the Oct-1 POU domain has no evident effect on DNA binding (Sturm & Herr, 1988). 3) Proteolysis of the POU domain reveals that the linker is readily accessible to proteases in solution (Botfield et al., 1992) and when bound to DNA (Aurora and Herr, 1992). 4) The amino acid sequence of the Oct-1 POU domain linker does not suggest any clear secondary structure propensity: It is primarily comprised of small polar or charged residues and also contains two prolines and two glycines. All the available evidence - including that fact that this region is disordered in the crystal - suggests that the linker is simply a flexible tether that provides a covalent connection between the POU-specific domain and the POU homeodomain.

Structure of the DNA

The 14 base pair DNA fragment forms a relatively uniform B-DNA helix, and there are no dramatic bends or distortions in the DNA. The average helical twist is 34.6° (10.4 residues per turn) and the average rise per base pair is 3.4 Å (Lavery & Sklenar, 1988). Although the overall structure of the DNA is clear at this stage, higher resolution data will be needed to define the sugar puckers and allow a base by base description of the DNA structure. DNA bending studies had suggested that the Oct-1 POU domain bends the DNA by about 37° (Verrijzer et al., 1991), and we see a net bend that is generally consistent with these studies. [Although the precise number depends on exactly how the helical axes are defined, our DNA has a net bend of approximately thirty degrees.]

Correlation with Biochemical Data

Our crystal structure of the Oct-1 POU domain—DNA complex is consistent with a large body of data about Oct-1—DNA interactions. Footprinting and chemical modification experiments and binding site selections with the two separate subdomains have shown that the Oct-1 POU-specific domain recognizes the 5' half of the octamer motif and that the POU homeodomain recognizes the 3' half (Verrijzer et al., 1990; Verrijzer et al., 1992), as seen in our crystal structure. Other chemical modification and mutagenesis studies have suggested that Oct-1—DNA contacts can extend across a 14 base pair region (Baumruker et al., 1988), which is exactly what we see in our complex. Finally, we note that the base contacts made by the POU domain are consistent with the optimal binding site selected for Oct-1 (Verrijzer et al., 1992). Multiple rounds of selection identified the canonical octamer sequence as the best binding site for Oct-1, but revealed an equal preference for adenine or thymine at position 5. In our complex, the adenine N3 (which is exposed in the minor groove) accepts a hydrogen bond from the side chain of Arg-5 of the POU homeodomain. As pointed out by Seeman et al.

(1976), the O2 of a thymine would have similar hydrogen bonding characteristics, and thus it is not surprising that either thymine or adenine is acceptable at position 5. Adenine is preferred at this position in naturally occurring octamer sites in cellular promoters and there may be additional constraints for the in vivo functioning of the octamer element.

Comparisons with the Phage Repressors and with Other Homeodomains Give New Perspectives on DNA Recognition

Detailed structural comparisons with other complexes give important new information about recognition and help us resolve several long-standing questions about these HTH proteins.

It is remarkable that the structure and key DNA contacts of the "λ-like" domain are so closely conserved between a set of bacteriophage repressors and the eukaryotic POU proteins, illustrating the utility and adaptability of this HTH-containing domain. It is also interesting to see how insertions are accommodated in the center of an HTH motif. As first seen in the NMR structures of the POU-specific domain (Assa-Munt et al., 1993; Dekker et al., 1993), the first helix of the HTH unit and the turn between the helices are each three residues longer than in the phage proteins. The insertion of six amino acids is readily accommodated in the complex because this region is directed away from the rest of the protein and from the DNA . Even longer insertions into this region are present in the POU domains of unc-86, Brn-3, and twin of I-POU (Finney et al., 1988; He et al., 1989, Treacy et al., 1992) and recent studies of other regulatory proteins have also shown that insertions can be accommodated in the middle of the HTH motif (reviewed in Brennan, 1993).

The striking similarity of the glutamine-DNA interactions in the POU domain—DNA complex and the phage repressor—DNA complexes is consistent

with the idea that conserved residues may have conserved roles when they appear at corresponding positions of a particular DNA-binding motif. (Although there is no general code for DNA sequence recognition, there may be "context dependent codes" for particular DNA-binding motifs.) As suggested by Pabo et al. (1990), it appears that the hydrogen bonding network made by the glutamine residues in this particular family of DNA-binding proteins is especially important for recognition.

The structure of the POU complex also has important implications for our understanding of homeodomain-DNA interactions. This structure is the first in which the N-terminal arm of a homeodomain has been studied in the context of a larger polypeptide, instead of being artificially located at the N-terminus of the molecule. Our structure confirms that the N-terminal arm makes critical contacts in the minor groove even in the context of a larger polypeptide (It seems ironic that the preceding residues from the linker region are disordered in our complex and it will be interesting to see whether this occurs in other homeodomain—DNA complexes). The importance of the N-terminal arm is highlighted by the fact that Ser-7, a residue in the arm of Oct-1, is a target for the cell-cycle regulation of DNA binding (Segil et al., 1991): Oct-1 is phosphorylated as cells complete DNA synthesis and enter mitosis (Roberts et al., 1991) and this phosphorylated form of the protein does not bind an H2B octamer (Segil et al., 1991). The phosphate is removed and DNA binding activity is restored as cells exit mitosis. In the complex, we find that Ser-7 is about 3.6 Å from a phosphodiester oxygen of adenine 2'. It seems clear that phosphorylation of Ser-7 would result in very unfavorable electrostatic interactions, and thus it is not surprising that phosphorylation would prevent DNA binding to this site. Additional evidence for the importance of the arm is provided by the Drosophila I-POU protein. I-POU lacks conserved basic residues at positions 3 and 4 of the

arm and cannot bind DNA (Treacy et al., 1991). Replacement of these residues with arginine and lysine in the alternatively spliced form (twin of I-POU) restores its DNA-binding properties (Treacy et al., 1992).

Our structure also provides new information about the role of Cys-50 in the POU homeodomains. Genetic and biochemical studies of homeodomains have shown that the residue at position 50 contacts bases on the 3' side of the TAAT site and plays an important role in determining the differential DNA binding affinity of homeodomain proteins (Hanes & Brent, 1989; Treisman et al., 1989). POU homeodomains always have a cysteine at that position and we find that Cys-50 is facing the major groove and interacts with bases just to the 3' side of the AAAT site. It is important to note, however, that these thymines (at positions 9' and 10') lie just outside of the standard octamer site. In fact, in vitro selection of optimized sites shows only a slight preference for thymine or adenine at position 9' and actually shows a slight preference for adenine at position 10' (Verrijzer et al., 1992). In addition, Ingraham et al. (1990) have shown that changing the Cys-50 in the Pit-1 POU domain to glutamine has little effect on DNA-binding affinity. Thus, our structure is consistent with the available biochemical data but it is still not clear why Cys-50 is strictly conserved in the POU homeodomains.

Understanding Cooperativity in POU domain-DNA Interactions

Biochemical studies have given important information about the DNA-binding activities of the POU homeodomain, the POU-specific domain and the intact POU domain. It has been shown that the isolated Oct-1 homeodomain binds DNA reasonably well but that the isolated POU-specific domain binds DNA quite poorly. Additionally, binding site selection studies indicate that Oct-1 prefers an arrangement in which the POU-specific binding site (ATGCAAAT)

and the POU homeodomain binding site (ATGCAAAT) are juxtaposed (Verrijzer et al., 1992). DNA binding competition experiments also support this conclusion. Changing the subsite spacing by one base pair (ATGCCAAAT) reduces the association constant about 10-fold and changing the subsite spacing by two base pairs (ATGCACAAAT) reduces the association constant about 30-fold (discussed in Chapter 3). How does the crystal structure of the Oct-1 complex let us understand and rationalize these binding data?

Perhaps the most surprising features of the Oct-1—DNA complex are that the linker region appears to be flexible and that there are no protein-protein contacts between the domains. We note that having a covalent connection between these DNA-binding domains will provide a way of enhancing binding affinity and specificity. Thus we can compare and contrast the POU complex with 1) the λ repressor dimer (where two identical DNA-binding motifs are held together by noncovalent interactions) or with 2) the TFIIIA family of zinc fingers (where related structural modules are covalently linked to enhance affinity and specificity) (Pabo & Sauer, 1992). Having a loose covalent connection between two DNA-binding domains illustrates another way that evolution has solved the problem of "designing" site-specific DNA-binding proteins and it appears that adding the POU-specific domain supplements the limited sequence specificity of the homeodomain (reviewed in Laughon, 1991). We do not know precisely how flexible the linker region is (we cannot rule out weak structural preferences that could play some role in positioning the domains) but even a completely flexible linker could enhance binding through a "chelate effect" (see Jencks, 1981): Whenever one domain is bound to its subsite, the linker will maintain a high local concentration of the "free" domain and thus facilitate binding or rebinding to a neighboring subsite.

Although our crystal structure shows that the two domains are structurally independent, they do make overlapping phosphate contacts near the center of the octamer site and this could provide another type of thermodynamic coupling in the binding process. In this region near the center of the site, the phosphates of base pairs 5 and 6 (in the "AAAT half" of the octamer sequence) are contacted by both the POU-specific domain and the POU homeodomain (Figure 9). [The arm of the homeodomain makes contacts from the minor groove while the POU-specific domain makes contacts from the major groove.] Thus, even though there are no direct protein-protein contacts between the domains, they may interact through subtle alterations in the structure or flexibility of the DNA. Such interactions may explain the preferred spacing of the subsites that is seen in the octamer sequence and the possibility of such coupling - through the DNA - is consistent with biochemical studies which suggest that the POU-specific domain and the POU homeodomain influence one another's DNA-binding specificity (Aurora & Herr, 1992; Verrijzer et al., 1992).

There are several other interesting implications of having 1) a flexible linker between the domains and 2) a complex that does not rely on protein-protein interactions between the two domains. This type of arrangement (which allows the two domains to bind on opposite sides of the DNA) leaves the surface of the POU-specific domain and the surface of the POU homeodomain fully accessible for interacting with other factors, such as VP16. These observations also help us understand how other POU domains can bind with other subsite spacings. It has been shown that the POU domain protein Brn-2 will bind to half-sites separated by 0, 2, or 3 base pairs, while Brn-3 binds exclusively to half-sites separated by 3 base-pairs (Li et al., 1993). Remarkably, the spacing preferences of these two domains can be swapped by swapping the N-terminal arm and helix 2

of the homeodomains of these proteins, further emphasizing the importance of the homeodomain arm in DNA recognition by the POU domain.

In summary, the structure of the Oct-1 POU domain—octamer complex and the biochemical studies of POU domain—DNA interactions raise important questions about cooperativity in protein-DNA recognition. Key issues involve: 1) the "chelate effect" or "local concentration" effects that result from having a covalent connection between the domains, 2) weak contacts or constraints that could (in principle) result from any residual structural preferences in the disordered linker region, and 3) thermodynamic coupling that may result from overlapping phosphate contacts near the center of the site. The relative importance of these terms may be different for different POU domains and different binding sites (such as for Brn-2 and Brn-3), but these are fundamental issues in protein-DNA recognition. Continued efforts to understand the coupling in the Oct-1-octamer interactions (discussed in Chapter 3) may have general implications in helping us understand how regulatory proteins can bind cooperatively to neighboring sites on double-helical DNA.

Conclusions

To summarize our key observations and the implications for understanding protein—DNA recognition, we note that:

The POU domain is a bipartite DNA-binding unit composed of two structurally independent DNA-binding domains tethered by an apparently flexible linker. The POU-specific domain makes its primary contacts with the ATGC sequence and the POU homeodomain makes its primary contacts with the AAAT sequence. These domains bind on opposite sides of the double helix, and there are no protein-protein contacts between them.

The POU-specific domain, which is structurally similar to the λ and 434 repressors, also docks on the DNA in a very similar way and some of the key DNA contacts are conserved. In particular, glutamines at the start of helix 2 and helix 3 form a conserved hydrogen bonding network in all of these complexes, indicating that this arrangement is particularly stable and is an essential feature in this family of HTH-containing domains.

The appearance of these "λ-like" domains in such divergent organisms (bacteriophage and eukaryotes) raises interesting evolutionary questions. Was this domain so successful that it has been conserved for the 1 - 2 billion years since these groups diverged? Was this domain transferred more recently by "horizontal transmission" between bacteriophage and eukaryotes (Bork & Doolittle, 1992)? Alternatively, are these examples of convergent evolution?

The overall structure and docking of the POU homeodomain is very similar to that seen in other homeodomain—DNA complexes, thus confirming and extending our general model for homeodomain—DNA interactions.

The structure gives new information about the role of Cys-50 but also raises intriguing new questions. We find that Cys-50, which is absolutely conserved among POU-type homeodomains, interacts with bases outside of the canonical octamer sequence; it is still not clear why this residue is invariant.

As observed in other homeodomain—DNA complexes, the N-terminal arm binds in the minor groove, near the 5' end of the AAAT site. This result is important since the POU domain shows the N-terminal arm within the context of a larger polypeptide and our structure confirms that the N-terminal arm makes minor groove contacts.

The linker connecting the POU-specific domain and the POU homeodomain varies considerably in length and amino acid sequence among various POU domain proteins. Our crystal structure - and the available

biochemical data - suggest that this linker region is very flexible. Using covalent connections between such distinct domains reveals an interesting and previously uncharacterized type of site-specific DNA-binding protein. Linking the domains in this way will ensure a high local concentration (whenever one domain is bound) and this should enhance binding through a chelate effect. Weak structural preferences in this linker region also could help to orient the domains as they bind to the subsites.

Although the POU-specific domain and the POU homeodomain make their primary contacts on separate halves of the octamer site, binding and selection data suggest that there is some thermodynamic coupling. In this context, the overlapping phosphate contacts that we see near the center of the octamer site may be very important. Many regulatory processes involve cooperative binding to neighboring sites or subsites on the DNA, and the structure of the Oct-1 complex provides a new basis for thinking about these interactions.

Given these novel features of the POU domain - and biochemical studies which indicate that other POU domains have distinct subsite spacing preferences - there are still many important questions to explore. The structure of the Oct-1—DNA complex should provide a useful basis for addressing these questions and for planning further thermodynamic, structural and biological studies of the POU domain.

Experimental Procedures

Protein and DNA Purification

The Oct-1 POU domain was prepared as a C-terminal fusion with glutathione-S-transferase. The chimeric protein was expressed, purified on glutathione-agarose (Sigma), and cleaved as described in Aurora and Herr (1992). The POU domain was further purified on a DNA cellulose column (Sigma) as follows: Protein was loaded to a 20 ml column in 0.05 M Tris pH 8.0, 0.15 M NaCl, 1 mM EDTA and 5 mM dithiolthreitol (DTT). The column was washed with 2 column volumes of 0.05 M Hepes (pH 7.5), 0.05 M NaCl, 0.01M DTT, followed by approximately 2 column volumes of the same buffer with 0.15 M NaCl. Protein was eluted from the column using this same buffer with 0.25 M NaCl. The purified protein was dialyzed extensively against double-distilled water at 4° C in 3500 molecular weight cut-off dialysis tubing and concentrated to approximately 10 mg/ml by lyophilization. The protein was not allowed to dry completely during lyophilization.

DNA duplexes used for crystallization were synthesized on an Applied Biosystems DNA/RNA Synthesizer 392. Individual strands were purified with the trityl group on by preparative reverse phase HPLC on a Vydac C_4 column using an acetonitrile gradient in 50 mM triethylammonium acetate (pH 6.5). The trityl group was cleaved by treatment with 1.1% trifluoroacetic acid for 10 minutes and the solution was immediately neutralized with triethylamine. Oligonucleotides were then dialyzed extensively against 10 mM triethylammonium bicarbonate (pH 7.0) and freeze-dried. The detritylated oligonucleotides were purified a second time by reverse phase HPLC and dialyzed against triethylammonium bicarbonate. DNA strands were annealed

by heating at 90° for 10 minutes and cooling slowly to room temperature. DNA duplexes were stored as freeze-dried aliquots.

Crystallization and Structure Determination

Oct-1 POU domain at 10 mg/ml in 10 mM DTT was combined with 2 equivalents of lyophilized binding site and incubated for approximately 0.5 hr at room temperature. Cocrystals were grown at room temperature by the hanging drop vapor diffusion method. One microliter of the complex was mixed on a coverslip with an equal volume of 30-34% PEG 3350, 0.10-0.15 M ammonium sulfate, and 0.05 M citrate pH 5.4, and was sealed over a well containing the same buffer. Crystals appeared in 1-3 days and grew to their maximum size in about 7 days. We confirmed that the POU domain in the crystals was intact by showing that protein from a dissolved crystal migrates on a denaturing gel at the position expected for the entire POU domain. The crystals form in space group $C222_1$, with cell dimensions a=97.5 Å, b=89.8 Å and c=80.0 Å. The diffraction of these crystals is markedly anisotropic: They diffract strongly along the DNA axis, but diffract less well in other directions. The current data set is weak beyond 3.0 Å.

Isomorphous derivatives were prepared by substituting 5-iodouracil for thymine in the DNA oligo. Cocrystals were grown with 1) substitutions at positions 9' and 10' 2) substitutions at positions 7' and 9' and with 3) a single substitution at position 5'.

Native and derivative diffraction data sets were collected on single crystals at room temperature on an R-AXIS imaging plate system (Table 1). Data were processed using DENZO and SCALEPACK (Z. Otwinowski, Yale University). Derivative structure factors were then local scaled to the native structure factors using DSCALEAD (Rould, 1989). Isomorphous difference Patterson maps (calculated with PROTEIN, Steigemann, 1982) clearly showed the

positions of the iodine atoms. These positions and occupancies were refined against centric reflections with REFINE from the CCP4 package (S.E.R.C. [U.K.] Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography, distributed from Daresbury Laboratory, Warrington, UK). Heavy atom parameters of each derivative were further refined with PHARE (CCP4) using cross-phased refinement, in which a single derivative is refined using native phases determined only from the other two derivatives. Phases were then generated with all three derivatives using both the isomorphous and anomalous differences, and an initial 3.0 Å MIR electron density map was calculated. This map was subject to one round of solvent-flattening (Wang, 1985) and heavy atom parameters were refined against the solvent-flattened phases (Rould et al., 1992). MIR phases were re-calculated using the refined heavy atom parameters and a final solvent-flattened electron density map was calculated. The final phases (from PHARE) have a mean figure of merit of 0.70 from 20.0 to 3.0 Å.

The electron density was very good for both the protein domains and for the DNA; however, the polypeptide linker between the domains could not be traced. Using the interactive graphics program FRODO (Jones, 1978), an initial model of the DNA in the complex was built using base pairs taken from a model of idealized B-form DNA. The positions of the iodine atoms were then used to properly align the sequence. The POU homeodomain electron density was readily fit with a model of the engrailed homeodomain backbone. The NMR structure of the POU-specific domain (Assa-Munt et al., 1993) was used as a starting model for the other domain. Each of the four helices in this domain were positioned separately and the loops between helices 2 and 3 and between helices 3 and 4 were rebuilt. When constructing the initial model, only side chains with

very clear electron density were included in order to minimize model bias during refinement.

The model was subject to multiple rounds of positional refinement using XPLOR (Brünger, 1992). After each round, side chains were added to the model as they became clear in $2F_O$-F_C maps. Positional refinement and model adjustment were then continued. Rebuilding was aided by comparing the original MIR electron density map with models from different simulated annealing runs (Brünger et al., 1987). In later stages of refinement, an overall anisotropic B correction was applied to the observed structure factors and tightly restrained individual B-factors were applied to each atom. Local scaling of the observed and calculated structure factors (using a minimum neighborhood of 100 reflections and excluding the reflection being scaled) was also done to correct for absorption during data collection. When the model was complete, it was checked against simulated annealing omit maps (Brünger, 1992). Refinement of the model with TNT (Tronrud et al., 1987) gave better geometry for the DNA, while refinement with XPLOR [using the protein bond and angle dictionary of Engh and Huber (1991)] gave better geometry for the protein. Thus, the final model is composed of TNT-refined DNA and XPLOR-refined protein. (In each case, the whole model was refined, but only the relevant portion was used for the final model.) Our final model includes residues 5-75 of the POU-specific domain and residues 2-61 of the POU-homeodomain. The polypeptide linker could not be traced in the original MIR map or in the final $2F_O$-F_C maps. The final R-factor is 23.7% for all data from 20-3.0 Å with excellent geometry: the rms deviation for bond lengths is 0.006 Å for the protein and 0.013 Å for the DNA; the rms deviation for bond angles is 1.12° for the protein and 2.50° for the DNA.

Acknowledgments

We thank Nuria Assa-Munt, Russell Mortishire-Smith, and Peter Wright (Scripps Research Institute) for the coordinates of the NMR structure of the POU-specific domain prior to publication; Michele Cleary (Cold Spring Harbor Laboratory) for invaluable help with DNA binding competition experiments; Nikola Pavletich (Memorial Sloan-Kettering Cancer Center) for advice during early stages of this project; Brigitte Raumann (Massachusetts Institute of Technology) for help during rebuilding and refinement; and members of the Pabo Lab for helpful discussions and for critical reading of the manuscript.

Note

In the time since this manuscript was published, we have noticed an additional protein-DNA interaction in the Oct-1 POU domain—octamer complex. Leu-55 of the POU-specific domain is in van der Waals' contact with the C5 methyl group of Thy-5'. The implications of this observation will be discussed in Chapter 3.

References

Aggarwal, A. K., Rogers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science *242*, 899-907.

Assa-Munt, N., Mortishire-Smith, R. J., Aurora, R., Herr, W., and Wright, P. E. (1993). The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. Cell *73*, 193-205.

Aurora, R., and Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. Mol. Cell. Biol. *12*, 455-467.

Baumruker, T., Sturm, R., and Herr, W. (1988). OBP100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. Genes Dev. *2*, 1400-1413.

Beamer, L. J., and Pabo, C. O. (1992). Refined 1.8 Å crystal structure of the l repressor-operator complex. J. Mol. Biol. *227*, 177-196.

Bodner, M., Castrillo, J., Theill, L. E., Deerinck, T., Ellisman, M., and Karin, M. (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell *55*, 505-518.

Bork, P., and Doolittle, R. F. (1992). Proposed acquisition of an animal protein by bacteria. Proc. Natl. Acad. Sci. USA *89*, 8990-8994.

Botfield, M. C., Jancso, A., and Weiss, M. A. (1992). Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. Biochemistry *31*, 5841-5848.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell *74*, 773-776.

Brünger, A. T. (1992). XPLOR v3.1 Manual. Yale Univ. Press, New Haven, CT.

Brünger, A T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. Science *235*, 458-460.

Cleary, M. A., and Herr, W. (1995). Mechanisms for flexibility in DNA sequence recognition and VP16-induced complex formation by the Oct-1 POU domain. Mol. Cell. Biol. *15*, 20090-2100.

Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., van der Vleit, P. C., and Kaptein, R. (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. Nature *362*, 852-855.

Engh, R. R., and Huber, R. (1991). Accurate bond and angle parameters for x-ray protein-structure refinement. Acta Cryst *A47*, 392-400.

Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The C. elegans cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors. Cell *55*, 757-769.

Gerster, T. and Roeder, R. G. (1988). A herpesvirus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. Proc. Natl. Acad. Sci. USA *85*, 6347-6351.

Hanes, S. D., and Brent, R. (1989). DNA specificity of the Bicoid activator protein is determined by the homeodomain recognition helix residue 9. Cell *57*, 1275-1283.

He, X., Treacy, M. N., Simmons, D. M., Ingraham, H. A., Swanson, L. W., and Rosenfeld, M. G., (1989). Expression of a large family of POU-domain regulatory genes in mammalian development. Nature *340*, 35-42.

Herr, W. (1992). Oct-1 and Oct-2: differential transcriptional regulation by proteins that bind to the same DNA sequence. In Transcriptional Regulation, S. McKnight and K. Yamamoto, eds. (Cold Spring Harbor Laboratory Press) pp. 1103-1135.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. Genes Dev. *2*, 1513-1516.

Ingraham, H. A., Chen, R., Mangalam, H. J., Elsholtz, H. P., Flynn, S. E., Lin, C. R., Simmons, D. M., Swanson, L., and Rosenfeld, M. G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell *55*, 519-529.

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1—Pit-1 interactions. Cell *61*, 1021-1033.

Jencks, W. P. (1981). On the attribution and additivity of binding energies. Proc. Natl. Acad. Sci. USA *78*, 4046-4050.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Cryst. *11*, 268-272.

Jordan, S. R., and Pabo, C. O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. Science *242*, 893-899.

Kemler, I., Schreiber, E., Müller, M. M., Matthias, P., and Schaffner, W. (1989). Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. EMBO J. *8*, 2001-2008.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579-590.

Kristie, T. M., LeBowitz, J. H., and Sharp, P. A. (1989). The octamer-binding proteins form multi-protein-DNA complexes with the HSV α-TIF regulatory protein. EMBO J. *8*, 4229-4238.

Laughon, A. (1991). DNA binding specificity of homeodomains. Biochemistry *30*, 11357-11367.

Lavery, R., and Sklenar, H. (1988). The definition of generalized helicoidal parameters and an axis of curvature for irregular nucleic acids. J. Biomol. Struct. Dynam. *6*, 63-91.

Li, P., He, X., Gerrero, M. R., Mok, M., Aggarwal, A., and Rosenfeld, M. G. (1993). Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. Genes Dev. *7*, 2483-2496.

Li, S., Crenshaw, E. B., Rawson, E. J., Simmons, D. M., Swanson, L. W., and Rosenfeld, M. G. (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene *pit-1*. Nature *347*, 528-533.

Mondragón, A., and Harrison, S. C. (1991). The phage 434 cro/ O_R1 complex at 2.5 Å resolution. J. Mol. Biol. *219*, 321-334.

Nelson, C., Albert, V. R., Elsholtz, H. P., Lu, I. -W., and Rosenfeld, M. G. (1988). Activation of cell-specific expression of rat growth hormone and prolactin genes by a common transcription factor. Science *239*, 1400-1405.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., and Wüthrich, K. (1990). Protein—DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. *9*, 3085-3092.

Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R., and Harrison, S. C., (1990). Conserved residues make similar contacts in two repressor-operator complexes. Science *247*, 1210-1213.

Pabo, C. O., and Sauer, R. T. (1992). Transcription factors: Structural families and principles of DNA recognition. Annu. Rev. Biochem. *61*, 1053-1095.

Poellinger, L., and Roeder, R. G. (1989). Octamer transcription factors 1 and 2 each bind to two different functional elements in the immunoglobulin heavy-chain promoter. Mol. Cell. Biol. *9*, 747-756.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell *59*, 573-580.

Qian, Y. Q., Otting, G., and Wüthrich, K. (1993). NMR[1] detection of hydration water in the intermolecular interface of a protein—DNA Complex. J. Am. Chem. Soc. *115*, 1190-1191.

Roberts, S. B., Segil, N., and Heintz, N. (1991). Differential phosphorylation of the transcription factor Oct-1 during the cell cycle. Science *253*, 1022-1026.

Rosenfeld, M. G. (1991). POU-domain transcription factors: pou-er-ful developmental regulators. Genes Dev. *5*, 897-907.

Rould, M. A., Perona, J. J., and Steitz, T. A. (1992). Improving multiple isomorphous replacement phasing by heavy-atom refinement using solvent-flattened phases. Acta Cryst. *A48*, 751-756.

Ruvkun, G., and Finney, M. (1991). Regulation of transcription and cell identity by POU domain proteins. Cell *64*, 475-478.

Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. USA *73*, 804-808.

Segil, N., Roberts, S. B., and Heintz, N. (1991). Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct-1 DNA binding activity. Science *254*, 1814-1816.

Steigemann, W. (1982). Program PROTEIN. MPI für Biochemie, Martinsried, Germany.

Stern, S., Tanaka, M., and Herr, W. (1989). The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature *341*, 624-630.

Sturm, R. A., and Herr, W. (1988). The POU domain is a bipartite DNA-binding structure. Nature *336*, 601-604.

Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst. *A43*, 489.

Treacy, M. N., He, X., and Rosenfeld, M. G. (1991). I-POU: a POU protein that inhibits neuron-specific gene activation. Nature *350*, 577-584.

Treacy, M. N., Neilson, L. I., Turner, E. E., He, X., and Rosenfeld, M. G. (1992). Twin of I-POU: a two amino acid difference in the I-POU homeodomain distinguishes an activator from an inhibitor of transcription. Cell *68*, 491-505.

Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell *59*, 553-562.

Verrijzer, C. P., Kal, A. J., and van der Vliet, P. C. (1990). The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. Genes Dev. *4*, 1964-1974.

Verrijzer, C. P., van Oosterhout, J. A. W. M, van Weperen, W. W., and van der Vleit, P. C. (1991). POU proteins bend DNA via the POU-specific domain. EMBO J. *11*, 3007-3014.

Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., van Leeuwen, H. C., Strating, M. J. J., and van der Vliet, P. C. (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. EMBO J. *11*, 4993-5003.

Wang, B. -C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90-112.

Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S. C. (1988). Structure of a phage 434 cro/DNA complex. Nature *335*, 789-795.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MATa2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-528.

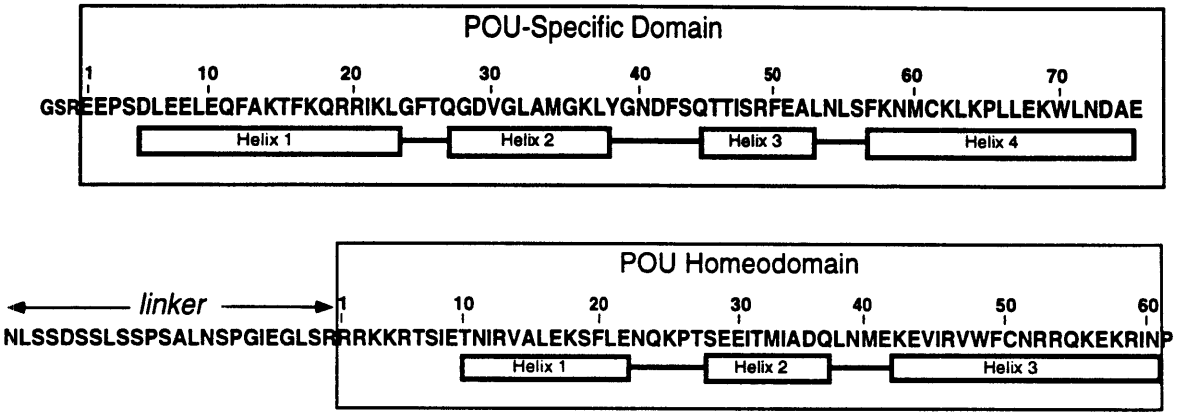

Figure 1. Sequence of the Oct-1 POU domain used in crystallization. The amino acid sequence is shown using the one letter amino acid code. The regions corresponding to the POU-specific domain and the POU homeodomain are boxed, and the linker segment is labeled. To facilitate comparison with other studies, the domains have been numbered independently. Shown below the sequence are the locations of the α helices seen in the crystal structure. Stippled helices in each domain indicate the HTH unit. The three residues preceding the POU-specific domain are extraneous residues from the expression construct.

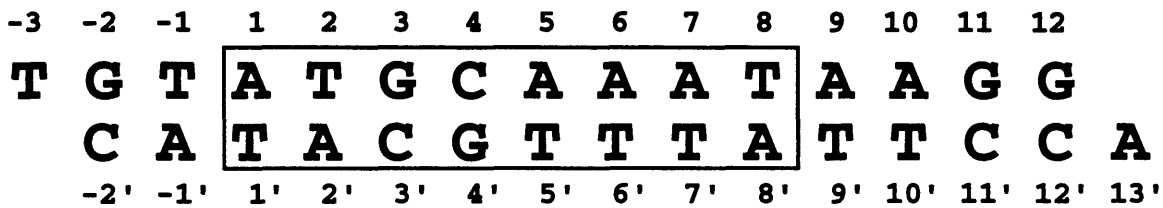

Figure 2. DNA duplex used for crystallization. The numbering of the nucleotides is centered on the octamer sequence, which is boxed. Base pairs -1 - 12 are from the human histone H2B promoter. The bases at the 5' ends mediate formation of a pseudocontinuous helix in the crystal.

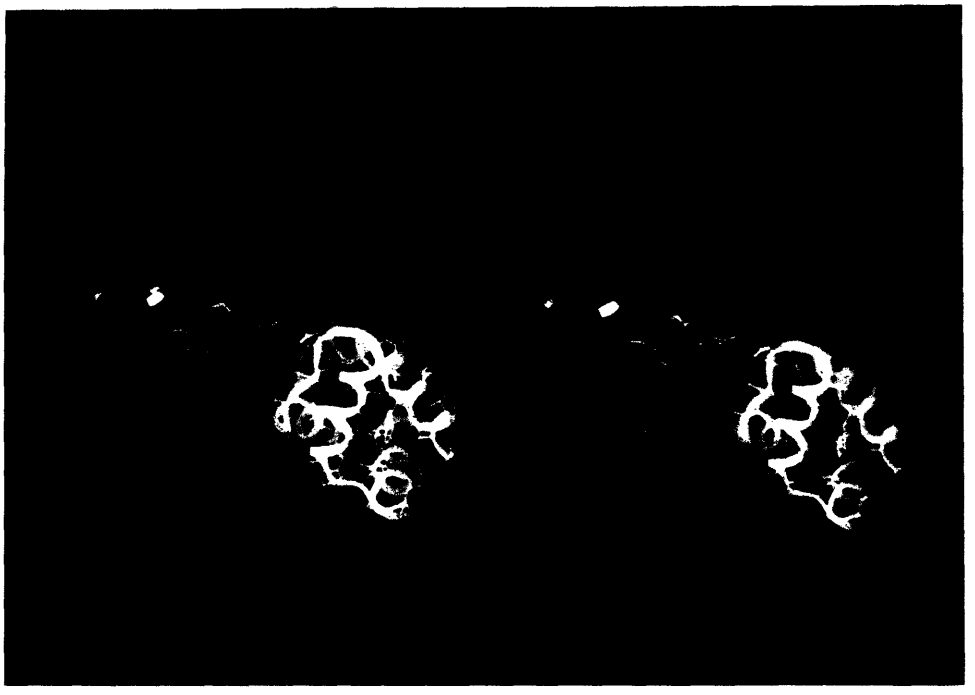

Figure 3. Overview of the Oct-1 POU Domain—DNA complex. The complex is shown as a stereo view, with ribbons drawn through the Cα's of the protein domains and through the phosphate backbone of the DNA strands. The POU homeodomain is red, and the POU-specific domain is yellow. The DNA is blue, with the octamer sequence highlighted in light blue.

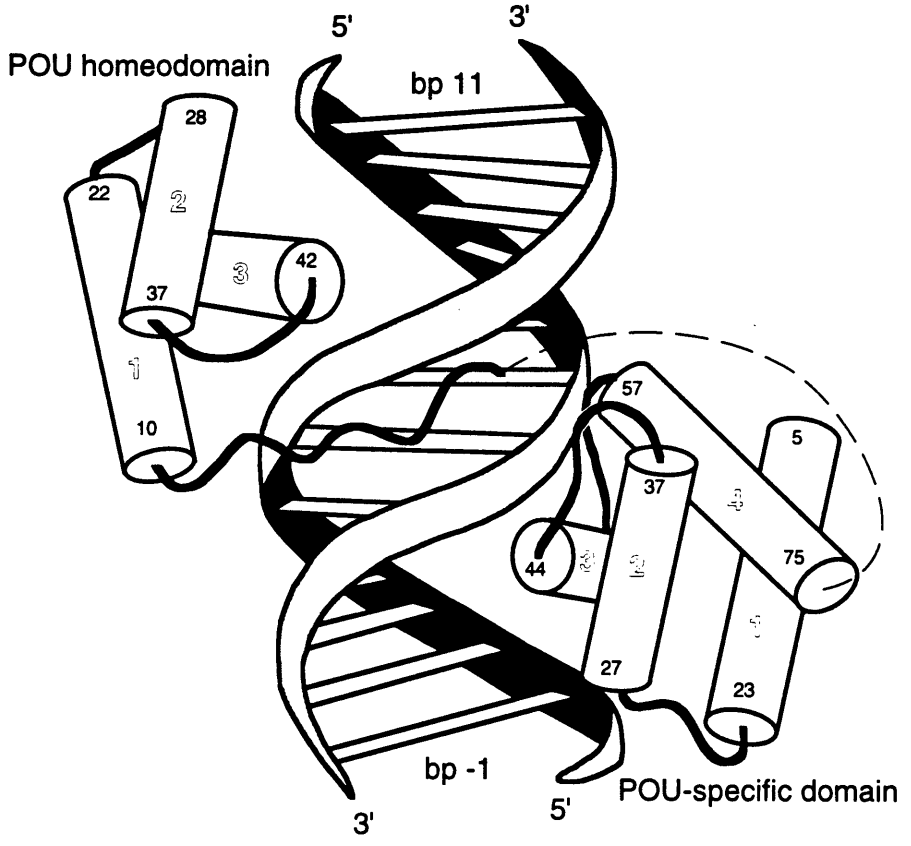

Figure 4. Overview of the POU Domain—DNA complex. Cylinders indicate the positions of the α helices, and the dotted line is used to emphasize that these domains are covalently connected (even though the linker is flexible and cannot be traced in the electron density maps).

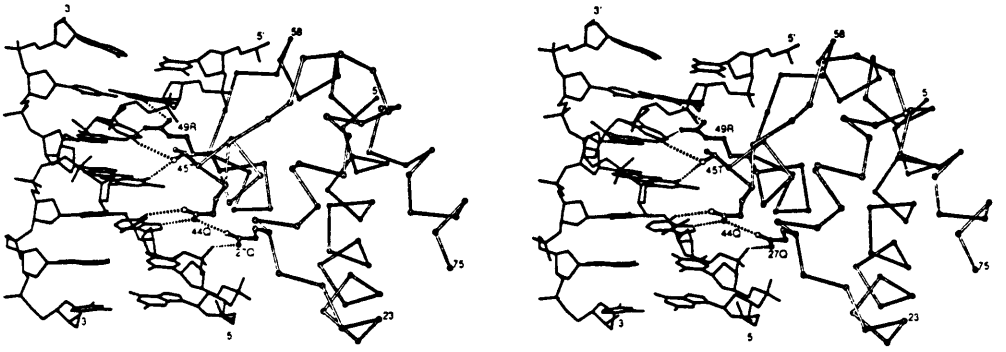

Figure 5. Stereo sketch showing the base contacts made by Gln-44, Thr-45 and Arg-49 of the POU-specific domain. These side chains are shown in bold. Gln-27, which interacts with Gln-44, also is shown.

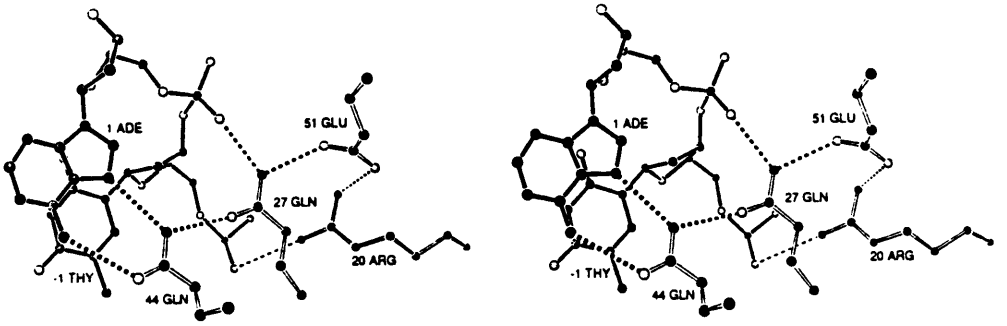

Figure 6. Stereo sketch of the complex hydrogen-bonding network made by Gln-44, Gln-27, Arg-20, Glu-51 and the DNA.

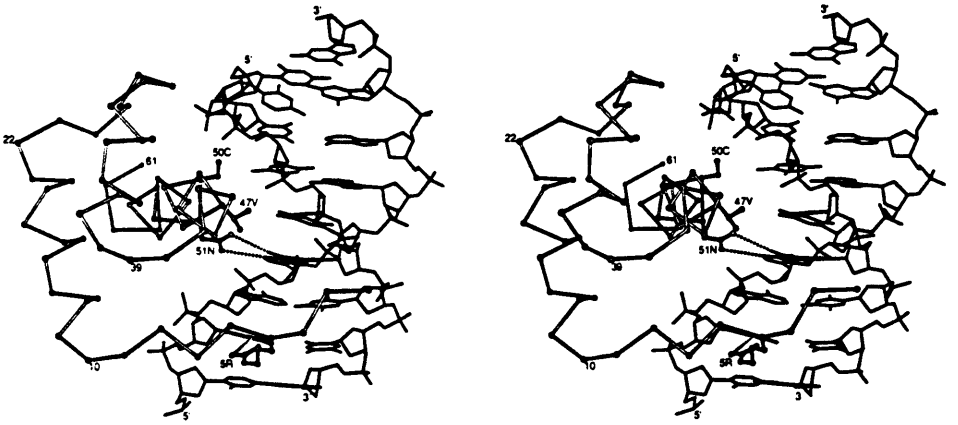

Figure 7. Stereo sketch showing the base contacts made by the POU homeodomain. Val-47, Cys-50 and Asn-51 in helix 3 make contacts in the major groove, and Arg-5 of the N-terminal arm makes a contact in the minor groove.

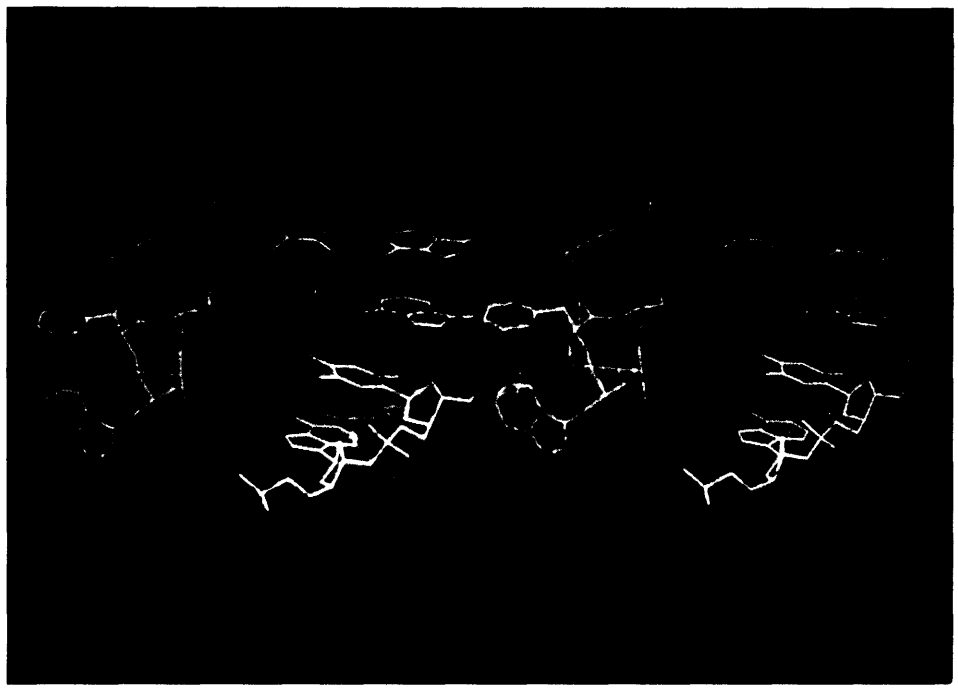

Figure 8. Stereo view of the solvent-flattened MIR electron density map at the POU homeodomain—DNA interface. The protein is shown in orange, the DNA is in yellow, and the electron density is blue. The map is contoured at 1.3 RMS above the average electron density.

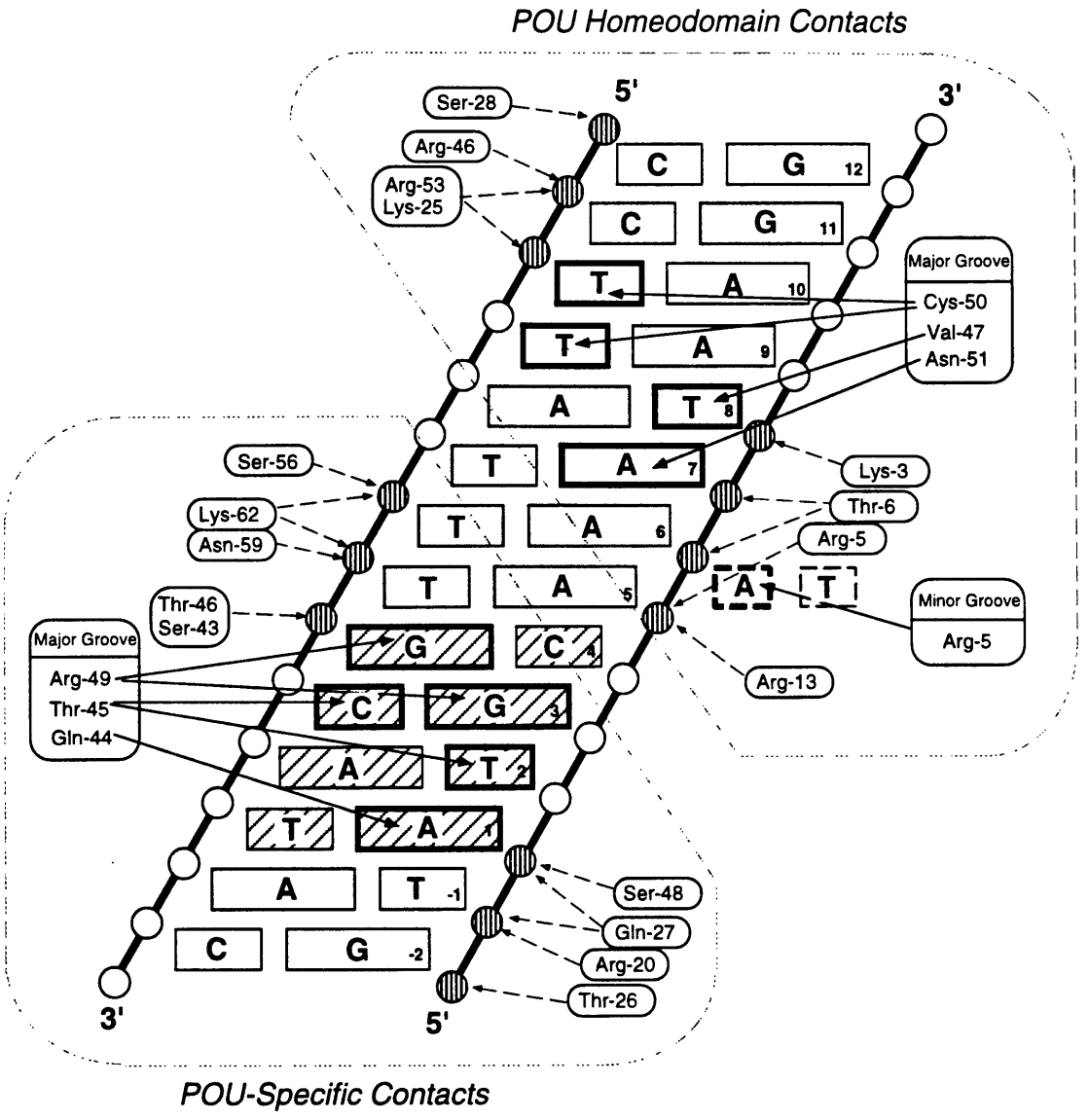

Figure 9. Sketch summarizing all the DNA contacts made by the POU domain. The DNA is represented as a cylindrical projection. Bases of the ATGC subsite are highlighted with hatched lines, and the bases of the AAAT subsite are stippled. Contacted bases are outlined in bold, and contacted phosphates are hatched. Most of the DNA contacts are made by amino acid side chains, but peptide NH groups make a few of the phosphate contacts. These include the following: the Lys-3 contact and one of the Thr-6 contacts in the POU homeodomain, and the Ser-43 contact and one of the Gln-27 contacts from the POU-specific domain.

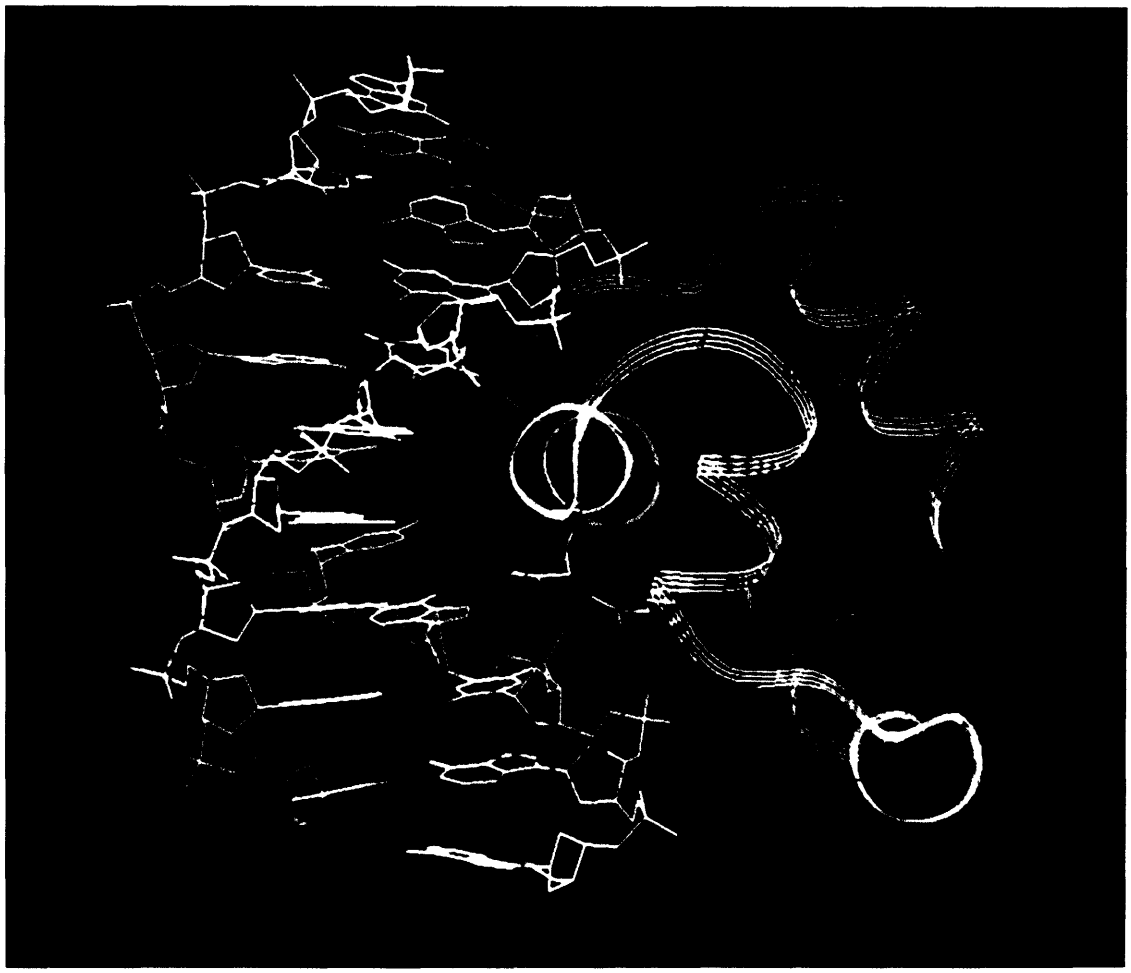

Figure 10. Superposition of the POU-specific domain bound to DNA with one λ repressor monomer bound to DNA (Beamer and Pabo, 1992). The POU-specific domain complex is green, and the λ repressor complex is white. The glutamines and the beginning of helices 2 and 3 of both proteins are also shown. The RMS deviation for the set of 54 corresponding Cα's is 1.9 Å.

Figure 11. Superposition of the POU homeodomain bound to DNA with the engrailed homeodomain bound to DNA (Kissinger et al., 1990). The POU homeodomain complex is yellow, and the engrailed complex is blue. Side chains at positions 47, 48, 49, 50, 51, and 53 of both proteins are shown. The RMS deviation for the Cα's of residues 9-58 is 0.99 Å.

	Native	dIU$^{5'}$	dIU$^{7'+9'}$	dIU$^{9'+10'}$
Resolution (Å)	3.0	3.0	3.0	3.0
Measured reflections	22,955	23,794	23,571	21,141
Unique reflections	6998	7034	7012	7018
Data coverage (%)	96	95	96	96
R_{sym}[a]	6.5	8.0	7.0	9.0
MIR Analysis				
Resolution limits (Å)		20.0-3.0	20.0-3.0	20.0-3.0
Phasing power[b]		1.59	2.65	2.14
Cullis R factor[c]		0.60	0.57	0.60
Mean figure of merit	0.70			
Refinement				
Resolution limits (Å)	20.0-3.0			
R factor (%)[d]	23.7			
Total number of atoms	1670			
Rms in B values (Å^2)[e]	1.91			
Protein geomety				
Rms in bond lengths (Å)		0.006		
Rms in bond angles (deg.)		1.20		
DNA geometry				
Rms in bond lengths (Å)		0.013		
Rms in bond angles (deg.)		2.50		

Designations for the derivative data sets indicate the base(s) at which 5-iodouracil (IU) was substituted for thymine.

[a] $\Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_{h,i}$, where I_h is the mean intensity of the i observations of reflection h.

[b] $[(F_{H(calc)})^2/(F_{PH(obs)}-F_{PH(calc)})^2]^{1/2}$

[c] $\Sigma||F_{der} \pm F_{nat}| - F_{H(calc)}|/\Sigma|F_{der} - F_{nat}|$ for centric reflections, where $F_{H(calc)}$ is the calculated heavy atom structure factor.

[d] $\Sigma|F_{(obs)} - F_{(calc)}|/\Sigma F_{(obs)}$ for all data

[e] For bonded atoms

Table 1. Statistics for Data and Derivatives

CHAPTER 3

Oct-1 POU Domain—DNA Interactions: Cooperative Binding of Isolated Subdomains and Effects of Covalent Linkage

Summary

Structural and biochemical studies of Oct-1 POU domain—DNA interactions have raised important questions about cooperativity and the role of the linker connecting the POU-specific domain and the POU homeodomain. To analyze these interactions, we have studied binding of the isolated domains. Surprisingly, we find that two unlinked peptides corresponding to the POU-specific domain and the POU homeodomain bind cooperatively to the octamer site and have a coupling energy of 1.6 kcal/mol. We suggest that overlapping DNA contacts near the center of the octamer site may be the source of this cooperativity, as there are no protein-protein contacts between the domains in the crystal structure of the Oct-1 POU domain—DNA complex. These studies have also allowed us to describe the contribution of the linker (present in the intact POU domain) in terms of an effective concentration. The broader implications for understanding cooperativity in protein-DNA recognition and gene regulation are discussed.

Introduction

The crystal structure of an Oct-1 POU domain—octamer (ATGCAAAT) complex has recently been determined (Klemm et al., 1994) and has raised interesting questions about cooperativity in protein-DNA interactions. In the crystal structure of the complex, the two globular domains are positioned on opposite sides of the double helix, and each makes specific base contacts to half of the octamer site: The POU-specific domain makes base contacts in the ATGCA subsite, and the POU homeodomain makes contacts in the AAAT subsite. The two domains make overlapping base and phosphate contacts near the center of the site; however, there are no protein-protein contacts between the two domains and the crystal structure suggests that the linker region is very flexible and would be long enough to allow binding with other spacings or other orientations of the domains.

In spite of the apparent structural autonomy of the two subdomains, biochemical evidence indicates that their binding is coupled. Thus, binding site selection experiments have shown that Oct-1 strongly prefers the wild-type arrangement of POU-specific domain and POU homeodomain subsites (Verrijzer et al., 1992). In agreement with this result, DNA binding competition experiments with the Oct-1 POU domain have shown that inserting 2 or 3 base pairs near the center of the octamer site significantly decreases the affinity for the Oct-1 POU domain (Klemm et al., 1994).

To determine why the Oct-1 POU-specific domain and POU homeodomain prefer to bind overlapping subsites, we have performed binding studies with purified peptides corresponding to the isolated domains. We find that the individual domains bind cooperatively to the octamer site,

and considering these results in conjunction with the cocrystal structure suggests that coupling may be mediated through overlapping DNA contacts near the center of the octamer site. Our data also allow us to describe the contribution of the linker that connects the POU-specific domain and the POU homeodomain in terms of an effective concentration. These results may have important implications for understanding cooperativity in protein-DNA recognition and for understanding the mechanisms of combinatorial control of gene expression.

Results

Individual Polypeptide Chains Containing the Oct-1 POU-Specific Domain and Oct-1 POU Homeodomain Form a Ternary Complex With the Octamer Site

Our strategy for analyzing cooperativity in POU domain—DNA interactions has involved studies of a ternary complex that contains the octamer site and two separate peptides corresponding to the POU-specific domain and the POU homeodomain. In initial studies, we used a gel mobility-shift assay to determine whether these two peptides can form a protein-DNA complex which mimics the intact POU domain—octamer complex. In a standard gel shift assay, the intact POU domain and the POU homeodomain each give a distinct mobility shift with the wild-type octamer probe (Figure 1, lanes 2 and 4). The isolated POU-specific domain does not shift this octamer probe under any conditions we have tested. However, under certain conditions (electrophoresis through an 18% gel at 4°C), we find that mixing the POU-specific domain and the POU homeodomain with the octamer probe results in formation of a complex with a mobility shift very similar to that of the

POU domain—octamer complex (Figure 1, lanes 3 and 4). The formation of this complex is specific for the wild-type octamer sequence. At the same protein concentrations, no supershifted band is formed on a probe in which key base pairs contacted by the POU-specific domain are changed from ATGCA to CGGCA (Figure 2, lane 7). The formation of this protein-DNA complex, requiring the presence of both polypeptides and giving a gel shift similar to that of the POU domain—DNA complex, suggests that the isolated domains can - under appropriate conditions - form a ternary complex with the octamer site.

DNase I footprinting provided further evidence that the complex we obtained with the isolated domains resembles the intact POU domain—DNA complex. The region of the octamer-containing DNA probe that is protected by the POU domain contains five strong DNase I bands (Figure 2, lane 2) while the region protected by the POU homeodomain includes four of these bands (Figure 2, lane 3). When the POU-specific domain and the POU homeodomain are combined, the footprinting pattern is extended to include all five bands that are characteristic of the intact POU domain footprint (Figure 2, lane 4). Thus, mixing the POU-specific domain and the POU homeodomain polypeptides can protect the same positions from DNase I cleavage that are protected by the intact POU domain. Both the footprinting and the gel shift results suggest that this ternary complex resembles a "linkerless" POU domain—DNA complex, and we have used this ternary complex for dissecting the cooperative interactions.

Cooperative Binding of the Oct-1 POU-Specific Domain and the Oct-1 POU Homeodomain

Further binding studies with the isolated domains show that the POU specific domain and the POU homeodomain can bind cooperatively even in the absence of the polypeptide linker. We measured the cooperativity in this system by comparing i) the affinity of the POU-specific domain for an unoccupied octamer site with ii) its affinity for an octamer site which has a POU homeodomain bound to the AAAT subsite (Figure 3). The difference in the free energies of binding under these two conditions reflects the cooperative interaction between the POU-specific domain and the POU homeodomain.

Comparing the DNase I protection pattern of the POU homeodomain with the combined protection pattern of the POU homeodomain and the POU-specific domain showed one clear band that could be used to monitor binding of the POU-specific domain. [This is the key difference between the footprint of the intact POU domain and the footprint of the isolated POU homeodomain.] To measure the K_d of the POU-specific domain, we quantitated the disappearance of this band as we increased the concentration of this domain. We measure a K_d of 2.2 ($\pm$ 0.14) x 10^{-5} M for the POU-specific domain—octamer complex, and this corresponds to a ΔG of -6.3 kcal/mol at room temperature (24°C) (Figures 4 and 5).

In parallel experiments, we measured binding of the POU-specific domain to a POU-homeodomain—octamer complex. In these experiments, the concentration of POU homeodomain was kept constant at 5 x 10^{-6} M and this concentration of POU homeodomain should give 97% saturation of the AAAT subsite. [The K_d we have measured for the POU homeodomain on the octamer sequence is 1.5 ($\pm$ 0.17) x 10^{-7} M.] Varying the concentration of the POU-specific domain and using DNase I footprinting to monitor binding shows that the K_d of the POU-specific domain under these conditions is 1.7 ($\pm$

0.88) x 10^{-6} M, corresponding to a ΔG of -7.9 kcal/mol (Figures 6 and 7). The higher affinity of the POU-specific domain for the POU homeodomain—octamer complex demonstrates that these subdomains bind cooperatively with a favorable interaction of -1.6 kcal/mol.

Cooperative Binding is Disrupted by Mutating the POU Homeodomain Binding Site or by Separating the Subsites

Control experiments confirm that the observed cooperativity requires both an intact homeodomain binding site and the normal arrangement of the ATGCA subsite recognized by the POU-specific domain and the AAAT subsite recognized by the POU homeodomain. One of the first control experiments was performed to confirm that occupancy of the POU homeodomain subsite was required for cooperative binding of the POU-specific domain. In this experiment, we repeated our cooperativity assay using a footprinting probe in which the POU homeodomain subsite and the flanking base pairs just outside of the octamer site (which are also contacted by the POU homeodomain in our crystal structure) were changed from AAAT a to AACGag (flanking bases in lowercase letters). No binding by the POU homeodomain could be detected on this probe, and we found that POU-specific binding to the probe was the same in the presence or absence of 5 x 10^{-6} M POU homeodomain (data not shown). As expected, cooperative binding of the POU-specific domain clearly requires that the POU homeodomain is bound at the AAAT subsite.

Previous studies had shown that the binding constant of the POU domain was reduced when two base pairs were inserted near the center of the octamer site to change the spacing of the ATGCA and AAAT subsites (Klemm et al., 1994). In parallel experiments using the isolated domains, we find that changing the spacing of the subsites abolishes the cooperative interactions. In

this set of experiments, we repeated the cooperativity assay (as illustrated in Figure 3) using a footprinting probe in which two base pairs were inserted between the octamer subsites. We find that this altered spacing does not affect the binding constants of the isolated domains, but we find that cooperativity is lost. With the altered subsite spacing, the K_d of the POU-specific domain is the same in the presence or absence of the POU homeodomain (data not shown). These experiments confirm that the wild-type arrangement of octamer subsites is required for cooperative binding of the POU-specific domain and the POU homeodomain.

Discussion

Cooperative Binding by the POU-Specific Domain and POU Homeodomain May be Mediated by Subtle Changes in DNA Structure

Our studies have focused on understanding the physical and structural basis for cooperativity in Oct-1 POU domain—DNA interactions. Our goal has been to understand why this protein prefers to bind the wild-type octamer site and why binding is less favorable if the subsites are moved apart by two base pairs. We also have been interested in understanding what contribution the linker makes to cooperative binding of the subdomains.

We have analyzed the mechanism of cooperativity using separate polypeptide chains for the POU-specific domain and the POU homeodomain, and we have shown that these domains bind cooperatively to the octamer site. Gel mobility-shift assays and DNase I footprinting experiments indicate that the isolated domains can mimic binding by the intact POU domain. Quantitative DNase I footprinting experiments have shown that there is a cooperative interaction energy of 1.6 kcal/mol. [Our studies compared POU-specific binding to the octamer site in the presence and absence of the

homeodomain, but in principle the complementary experiment could also be performed: I.e., binding of the POU homeodomain to the octamer site could be compared in the presence and absence of the POU-specific domain. This should give equivalent information about coupling but is technically more difficult because very high concentrations of POU-specific domain would be needed to ensure the stoichiometric formation of a POU-specific domain—octamer complex.]

We find cooperative binding of the isolated subdomains is dependent on the wild-type spacing of subsites: inserting two base pairs between the subsites abolishes the cooperative effect. This suggests that cooperative binding of the isolated domains may reflect the same energetic factors that cause the intact POU domain to prefer the wild-type octamer site. The qualitative dependence on the spacing is similar and the energies match reasonably well. Specifically, binding studies with the intact POU domain show that the binding energy is reduced by 2.1 kcal/mol when the subsites are separated by two base pairs. In studies with the isolated domains, we find that the 1.6 kcal/mol cooperative interaction is lost when the subsites are separated by two base pairs (data not shown). Thus, most of the binding energy that is lost when the intact POU domain binds to separated subsites may be due to the loss of cooperative interactions. [Although the linker is quite long and presumably very flexible, it is possible that there is some strain in the linker polypeptide with this spacing, and this also could contribute to the loss of binding energy.]

Evaluating our binding data in light of the known crystal structure of the Oct-1 POU domain—DNA complex suggests that the cooperativity may be mediated by overlapping DNA contacts near the center of the octamer site. Specifically, the crystal structure of the Oct-1 POU domain-DNA complex

shows that there are no protein-protein contacts between the domains, and thus we have focused on ways that the cooperativity could be mediated via the DNA. In this regard, we note that there are a set of overlapping contacts at the center of the octamer site. The A/T base pair at the fifth position of the octamer site (ATGCAAAT) is contacted in the major groove by a leucine from the POU-specific domain and in the minor groove by an arginine from the POU homeodomain. In addition, at positions 5 and 6 (ATGCAAAT) phophodiester oxygens on one DNA strand are recognized by the POU-specific domain while phosphodiester oxygens on the other DNA strand are recognized by the POU homeodomain. Given these extensive contacts, it certainly seems plausible that binding of the homeodomain could affect the conformation or flexibility of the DNA in this region and that subtle changes might facilitate binding of the POU-specific domain. Although we cannot precisely identify which regions of the DNA are involved, it does appear that cooperative binding by the POU-specific domain and the POU homeodomain is mediated via the DNA.

Contribution of the Linker to Cooperative Binding

In the intact POU domain, the linker facilitates binding by a chelate effect that maintains a high local concentration of the domains. Covalent linkage between the POU-specific domain and the POU homeodomain reduces the overall binding reaction from a three-body problem to a two-body problem, and this reduces the entropic cost of complex formation. Now that we can separate out the other cooperative interactions, our new data now allow us to quantitatively evaluate the contribution of the linker. Knowing the binding energies of the intact POU domain and of the isolated subdomains allows us to describe the contribution of the linker in terms of an effective

concentration. In simplest terms, the effective concentration provides a way of describing the chelate effect. Whenever one domain of this bipartite protein is bound to DNA, the linker will tether the other domain to a nearby region of space and hold it - at a high local concentration - near the DNA. A value for the effective concentration in this system can be calculated as:

$$C_{eff} = \frac{(K_d)^*_{POU_S}(K_d)_{POU_{HD}}}{(K_d)_{POU}}$$

where $(K_d)^*_{POU_S}$ is the K_d of the POU-specific domain in the presence of the POU homeodomain (see Creighton, 1984, and Experimental Procedures for a further discussion of effective concentration). Substituting the values obtained in our experiments shows that the effective concentration of either domain in an intermediate complex (when one domain is bound to the DNA but the other is still free) is 3.6×10^{-3} M. Since this effective concentration is significantly larger than the binding constant of either domain, it appears that any intermediate complex will be thermodynamically unstable and can only be present as a transient intermediate during binding or release of the POU domain. The flexible linker clearly makes a significant contribution to binding.

DNA Binding by Other POU Domain Proteins

Our studies of cooperativity in Oct-1 POU domain—DNA interactions may provide a model system for analyzing cooperativity in other protein-DNA interactions, but the precise nature of these interactions will be specific to each system. In this regard, it is important to realize that there may be fundamental differences in cooperativity and subsite spacing with other POU

domain proteins. Studies of DNA recognition by the POU domain proteins Brn-2, Tst-1, Unc-86, and Brn-3 have shown that they prefer to bind sites in which i) the POU-specific subsite is inverted relative to its orientation in the octamer sequence (ATGCA -> TGCAT) and ii) the POU-specific domain and POU homeodomain subsites are separated by two base pairs (for Brn-2 and Tst-1) or three base pairs (for Unc-86 and Brn-3) (Xue et al., 1992; Li et al., 1993). [In fact, Brn-2 also is capable of binding to subsites spaced by 0 or 3 base pairs.] The different orientations and spacings of the subdomains in these other POU domain—DNA complexes presumably involve differences in the mechanisms of coupling between the subdomains. In each system, we expect that the covalent linkage between the POU-specific domain and the POU homeodomain will raise the effective concentration and stabilize binding, but the precise magnitude of this effect will depend on the length and sequence of the linker and the relative arrangement of the binding sites. Our studies suggest that cooperative binding of the isolated domains has provided another mechanism for the evolutionary diversification of the POU domains, and differences within the family may help us understand how the POU domain has adopted a range of roles in protein-DNA recognition and gene regulation.

Relevance to Other Studies of Protein-DNA Interactions

Studies of covalently linked DNA-binding domains involve a new level of complexity in understanding protein-DNA interactions, and our experimental strategies should be applicable to studies of other DNA-binding proteins that have two or more covalently linked domains. Corresponding studies of TFIIIA-like zinc fingers could be particularly interesting. The linkers clearly play an important role in recognition (isolated fingers do not

bind effectively (Frankel et al., 1987)), but it would be very interesting to see whether there was any cooperativity left in the absence of a linker. In analogy to the POU domain, features of TFIIIA-type zinc finger—DNA complexes hint that DNA binding by adjacent fingers might be thermodynamically and structurally coupled. Crystal structures of these complexes show that each finger makes base contacts in a short 3 - 4 base pair subsite, but there often are overlapping contacts with other subsites (Pavletich and Pabo, 1991; Pavletich and Pabo, 1993; Fairall et al., 1993). For instance, finger 4 of GLI contacts three phosphates in the finger 5 subsite (Pavletich and Pabo, 1993). In the case of the zinc fingers, there is also evidence for a hydrogen bond between neighboring fingers, usually involving an arginine in the α-helix of one finger and a carbonyl oxygen in the turn of a neighboring finger. Finally, an analysis of the DNA structure in zinc finger—DNA complexes reveals an enlarged major groove (Nekludova and Pabo, 1994) and thus subtle changes in DNA structure may be involved in binding. Given these possibilities, it would be very interesting to remove one linker from a zinc finger protein and see whether the terminal (nth) finger would bind cooperatively in the presence of a peptide containing the other fingers 1 through n-1. Measuring binding constants for these fragments would also give important information about the contribution of the linker to recognition.

Our data suggest that cooperativity in the Oct-1 POU domain—DNA interactions - seen in binding of the isolated domains - may be mediated by overlapping DNA contacts near the center of the octamer site and may involve subtle changes in the DNA structure. It is widely accepted that binding of a protein to DNA may affect the overall structure of its binding site (reviewed by Travers, 1992) but it has not been clear how these changes might affect the interactions of other proteins with neighboring sites. In some

complexes, such as the CAP—DNA complex (Schultz et al., 1991) and the TBP—DNA complex (Kim et al., 1993; Kim et al., 1993), the DNA double helix is drastically bent or distorted by the protein. However, in other protein-DNA complexes, the changes in DNA structure are less pronounced. In this report, we have shown that the POU subdomains can bind cooperatively to overlapping subsites - in the absence of protein-protein interactions - and we think this involves subtle, interdependent effects on the structure and/or flexibility of the DNA at the juncture of the subsites. This type of indirect interaction between proteins bound to neighboring DNA sites may be one mechanism by which multiple transcription factors interact cooperatively in the combinatorial control of transcription. While many regulatory interactions have been shown to involve direct protein-protein contacts, the contribution from subtle changes in DNA structure has been difficult to assess. The Oct-1 POU domain has provided us with an important model system for studying how DNA-mediated effects may lead to cooperativity in protein-DNA interactions.

Experimental Procedures

Protein Constructs

The POU domain, the POU homeodomain and the POU-specific domain from Oct-1 were each expressed as C-terminal fusions with glutathione S-transferase (GST). The POU domain and POU-specific domain constructs were gifts of Winship Herr (Cold Spring Harbor). Construction of the POU domain expression vector is described in Aurora and Herr (1992); the POU-specific vector is a derivative of this. The POU homeodomain expression vector was constructed by PCR amplifying the appropriate region from the POU domain construct and cloning the fragment into the BamHI and XhoI sites of pGEX-4T-2 (Pharmacia). In order to facilitate cleavage of the GST tag from the POU homeodomain, three glycine residues were included after the thrombin cleavage site (Guan and Dixon, 1991). Therefore, the resulting protein has the five additional N-terminal amino acids GSGGG. The construct was verified by dideoxy sequencing.

Soluble proteins were expressed, purified on glutathione-agarose (Sigma) and cleaved as described in Aurora and Herr (1992). The POU-specific domain was concentrated using Centriprep and Microcon concentrators (Amicon). The POU domain was further purified on a DNA-cellulose column and concentrated as described previously (Klemm et al., 1994). Further purification of the POU homeodomain was carried out using Sep-Pak C_{18} reversed phase cartridges (Millipore). Sep-Paks were primed with 5 ml acetonitrile plus 0.1% trifluoroacetic acid (TFA), then equilibrated with 5 ml 4M guanidine hydrochloride (GuHCl) in 25 mM Tris (pH 7.5). The Sep-Pak was loaded with 8 ml of glutathione-agarose—purified POU homeodomain in 4 M GuHCl, 25 mM Tris (pH 7.5), and was then washed with equilibration

buffer. Protein was eluted using steps of acetonitrile (with 0.1% TFA); homeodomain eluted at 40% acetonitrile and was lyophilized. All proteins were quantitated by absorbance readings at 280 nm using the following calculated molar extinction coefficients: POU domain: 12,900 M^{-1} cm^{-1}; POU homeodomain: 5810 M^{-1} cm^{-1}; POU-specific domain: 7090 M^{-1} cm^{-1}.

Gel Mobility-Shift Experiments

The sequence of the synthetic probe containing the wild-type octamer is GGGCGCTGT**ATGCAAAT**AAGGCGCCC (octamer sequence in bold); the bottom strand is complementary to this. The mutant probe is a derivative of the wild-type probe in which two base pairs of the octamer site (in the subsite recognized by the POU-specific domain) have been changed (ATGCAAAT -> CGGCAAAT) to disrupt binding of the POU-specific domain. Oligonucleotides used for these studies were synthesized on an Applied Biosystems Model 390 DNA synthesizer and purified on 8% or 10% polyacrylamide-7M urea gels. Double-stranded oligonucleotides were 5' end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase (New England Biolabs). Unincorporated nucleotides were removed using a G-25 Sephadex quick spin column (Boehringer Mannheim). Binding reactions were carried out in a volume of 12.5 µl in a buffer containing 8 mM Hepes (pH 7.8), 100 mM NaCl, 0.03% NP-40, 5 mM $MgCl_2$, 0.1 mg/ml BSA, 1% ficoll and the appropriate amount of protein. The DNA probe was present at concentrations of less than 5 pM. For measuring POU domain K_d's, reactions were incubated at room temperature for 30 minutes, then loaded to 10% 0.5x Tris-borate-EDTA (TBE) gels and electrophoresed at 150 V. The dissociation constants measured for the POU domain on the wild-type octamer and on an octamer in which the subsites have been shifted by 2 base pairs are as follows:

wild-type octamer: 7.1 (± 1.9) x 10^{-11} M; octamer with 2 base pair insert: 2.3 (± 0.37) x 10^{-9} M. The DNA probe used in the experiment with the separated subsites is analogous to the wild-type probe (above) but has the sequenceATGCACAAAT.

For supershift assays with the POU homeodomain and the POU-specific domain, reactions were incubated on ice for 30 minutes, then loaded to an 18% 0.5x TBE gel and electrophoresed at 300 V at 4°C. In order to detect a supershift, relatively high protein concentrations were necessary. The POU homeodomain concentration in this experiment was 5 x 10^{-6} M and the POU-specific domain concentration was 1 x 10^{-4} M.

Footprinting Probe Construction, Purification, and Labeling

Footprinting probes were prepared by using synthetic oligonucleotides cloned into the EcoRI and BamHI restriction sites of pBluescript KS+. The sequence of the top strand of the synthetic oligonucleotide is:
AATTCCTGATCAAGATCTGGTCACCCCATGGGCTAGCGCATGCCCAAGG
CTGT**ATGCAAAT**AAGGACGCGTTCGCGAGGGCCCG (octamer sequence shown in bold). The sequence of the bottom strand of the synthetic oligonucleotide is:
GATCCGGGCCCTCGCGAACGCGTCCTT**ATTTGCAT**ACAGCCTTGGGCATG
CGCTAGCCCATGGGGTGACCAGATCTTGATCAGG (octamer sequence shown in bold). The oligo for studies with separated subsites is similar, but has two base pairs inserted near the center of the octamer site:
ATGCACAAAT. The synthetic duplexes were designed such that the 5' and 3' ends were ready for cloning into the EcoRI/BamHI-cut plasmid. Constructs were confirmed by dideoxy sequencing. DNA probes for footprinting were prepared by digesting the appropriate plasmid with EcoRI and SacI; the

resulting fragments are 115 (wild-type octamer) or 117 (altered subsite spacing) base pairs in length. After restriction digestion, the fragments for footprinting were purified by electrophoresis through 2% low melting point agarose gels. Excised bands were digested with Gelase (Epicentre Technologies) according to the manufacturer's protocol. The probe was radioactively labeled by filling in the EcoRI 5' overhang using the Klenow enzyme (Boehringer Mannheim) in the presence of [α-^{32}P]dATP and [α-^{32}P]dTTP.

DNase I Footprinting

Binding reactions were carried out at room temperature in a 200 µl volume with a buffer containing 8 mM Hepes (pH 7.8), 100 mM NaCl, 1 mM dithiothreitol, 0.1mg/ml BSA, 0.03% NP-40, 5 mM $MgCl_2$, 1 mM $CaCl_2$ and 20,000 cpm of probe DNA. The concentration of probe DNA in the reactions was less than 0.1 nM. After 30 minutes incubation, 40 µl of 2.5 ng/ml DNase I (Worthington) was added and the reaction was left at room temperature for 1 minute. The reaction was stopped by adding 200 µl of a buffer containing 2.5 M NH_4COOH, 1 mg/ml sonicated salmon sperm DNA, and 20 mM EDTA. The DNase I-treated probe was purified by extraction with phenol and chloroform and then precipitated with ethanol in the presence of 20 µg glycogen. Samples were denatured in formaldehyde loading buffer by boiling for 5 minutes, then analyzed on gels containing 8% polyacrylamide and 7 M urea.

Quantitation of Dissociation Constants

All gels were quantitated using a Molecular Dynamics 492I Phosphor Imager. For quantitation of footprinting gels, fractional protection of a site (p_i) was calculated as:

$$p_i = 1 - \left(\frac{I_{site} / I_{std}}{I_{ref,site} / I_{ref,std}} \right)$$

where I_{site} is the integrated intensity within a defined region, I_{std} is the integrated intensity of a standard region on the DNA fragment, and $I_{ref,site}$ and $I_{ref,std}$ are analogous parameters derived from a reference lane with no protein present (Brenowitz et al., 1993). The fractional protection was assumed to represent the fraction (θ) of bound DNA. For gel shift experiments, θ was calculated as:

$$\theta = \frac{I_{bound}}{I_{free} + I_{bound}}$$

Equilibrium dissociation constants (K_d) were determined by linear regression using the Scatchard equation:

$$\frac{\theta}{[P]} = \frac{1}{K_d} - \frac{\theta}{K_d}$$

where [P] represents the free protein concentration. For K_d calculations, we used points for which $0.2 \leq \theta \leq 0.8$. Because the total DNA concentration used in our binding experiments was always well below the K_d, the free protein concentration and the total protein concentration are essentially equal. Reported K_d's were derived from the slope of the Scatchard plot and represent the mean of at least of three independent experiments.

Effective Concentration

We have used the term "effective concentration" to describe the contribution the flexible linker makes to cooperative binding of the POU subdomains. This provides a quantitative way of describing the chelate effect. In the calculation of effective concentration described in the text, we use the dissociation constant for the POU-specific domain that is observed in the

presence of the homeodomain, $(K_d)^*_{POU_S}$. Identical results would be obtained by calculating:

$$C_{eff} = \frac{(K_d)_{POU_S}(K_d)_{POU_{HD}} K_{cooperativity}}{(K_d)_{POU}}$$

where $K_{cooperativity} = e^{\left(\frac{-\Delta G_{cooperativity}}{RT}\right)}$.

Thinking of the effective concentration that we have calculated (3.6 mM) in terms of the volume accessible to a single molecule shows that this corresponds to having the free domain constrained to a spherical volume of radius 48.1 Å. A rigorous geometric and energetic interpretation of the effective concentration would be extremely difficult, but we note that this radius falls well within the distance that could be spanned by the 24 amino acid Oct-1 POU domain linker (which might cover 24 x ~3.5 Å = 84 Å in a fully extended linker).

Acknowledgments

We thank Winship Herr and Reggie Aurora for the Oct-1 POU domain and POU-specific domain expression plasmids; Tracy Smith for advice on DNase I footprinting; Ernest Fraenkel and Harvey Greisman for critical reading of the manuscript; and Robert Sauer for helpful discussions.

References

Aurora, R., and Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. Mol. Cell. Biol. *12*, 455-467.

Brenowitz, M., Senear, D., Jamison, E., and Dalma-Weiszhausz, D. (1993). Quantitative DNase I Footprinting. In Footprinting of Nucleic Acid—Protein Complexes, (ed. A.Revzin), pp. 1-43. Acedemic Press, San Diego, California.

Creighton, T. E. (1984). Proteins: Structures and Molecular Principles, pp. 360-365. W. H. Freeman and Company, New York.

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extensiton to the rules for zinc-finger/DNA recognition. Nature *366*, 483-487.

Frankel, A. D., Berg, J. M., and Pabo, C. O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. USA *84*, 4841-4845.

Guan, K., and Dixon, J. E. (1991). Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathineS-transferase. (1991). Anal. Biochem. *192*, 262-267.

Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.

Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Li, P., He, X., Gerrero, M. R., Mok, M., Aggarwal, A., and Rosenfeld, M. G. (1993). Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. Genes Dev. *7*, 2483-2496.

Nekludova, L., and Pabo, C. O. (1994). Distinctive DNA conformation with enlarged major groove is found in Zn-finger—DNA and other protein—DNA complexes. Proc. Natl. Acad. Sci. USA *91*, 6948-6952.

Pavletich, N. P. and Pabo, C. O. (1991). Zinc finger—DNA recognition: crystal structure of a zif268-NDA complex at 2.1 Å. Science *252*, 809-817.

Pavletich, N. P. and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701-1707.

Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. Science *253*, 1001-1007.

Travers, A. A. (1992). DNA Conformation and Configuration in Protein—DNA Complexes. Curr. Opin. Struct. Biol. *2*, 71-77.

Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., van Leeuwen, H. C., Strating, M. J. J., and van der Vliet, P. C. (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. EMBO J. *10*, 4993-5003.

Xue, D., Finney, M., Ruvkin, G., and Chalfie, M. (1992). Regulation of the Mec-3 gene by the C. elegans homeoproteins unc-86 and Mec-3. EMBO J. *11*, 4969-4979.

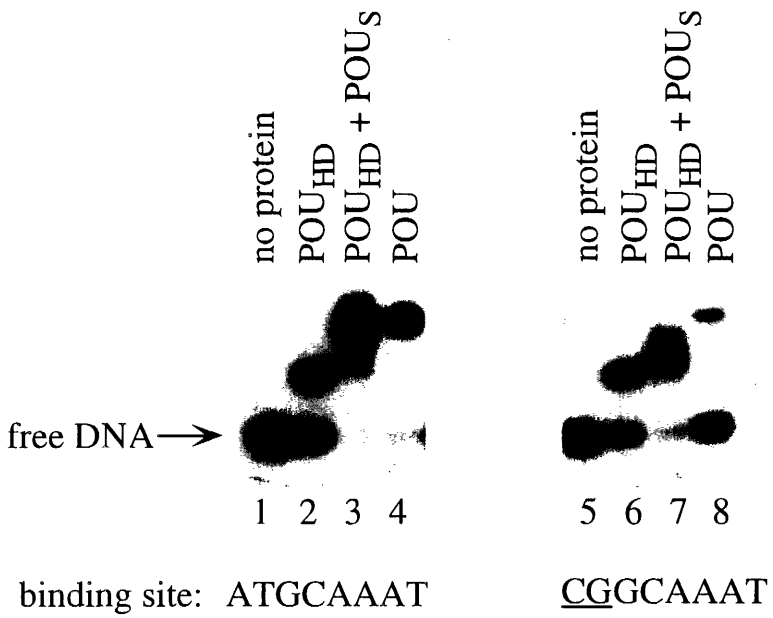

Figure 1. Gel mobility-shift assay showing that the POU-specific domain and the POU homeodomain form a ternary complex with the octamer site. The wild-type octamer probe (Lanes 1 - 4) is 26 base pairs in length and contains the sequence ATGCAAAT. In the mutant octamer probe (Lanes 5 - 8), two base pairs in the POU-specific subsite have been changed (ATGCA -> CGGCA).

In the experiment using the mutant (CGGCAAAT) probe, addition of the POU-specific domain to the POU homeodomain binding reaction does not produce a supershifted band, but the apparent affinity of the POU homeodomain for the DNA is increased (compare lanes 6 and 7). Given the very high concentration of protein used in these experiments, we suspect that the POU-specific domain may still weakly associate with the mutant site. We believe that this facilitates homeodomain binding but that the interactions of the POU-specific domain with the mutant half site are weak enough that this domain falls off during electrophoresis.

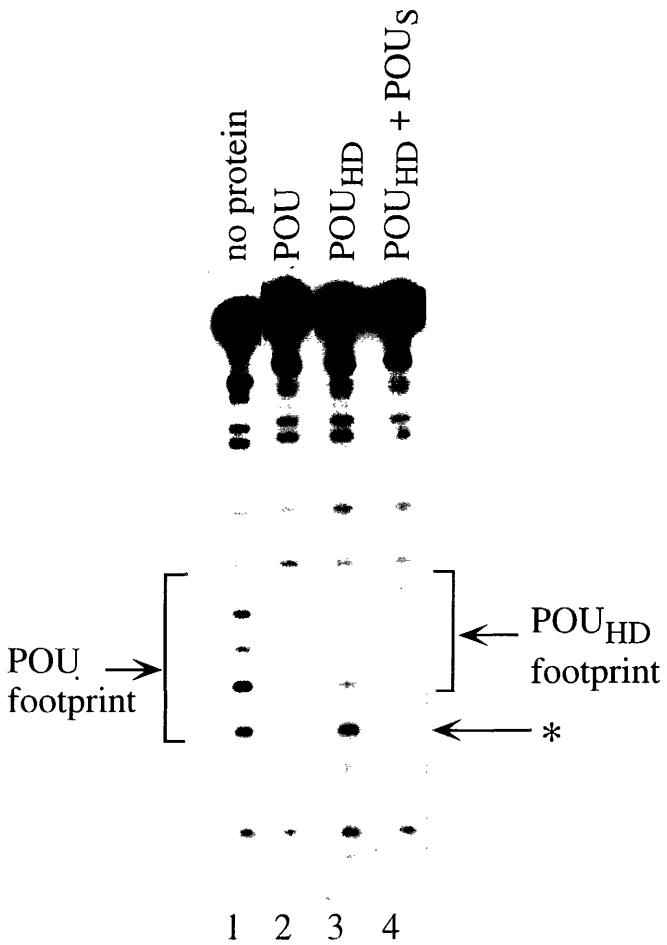

Figure 2. Comparison of the footprinting patterns of the POU domain, the POU homeodomain, and a mixture containing the POU homeodomain and the POU-specific domain. The DNA probe used in this and subsequent footprinting experiments is 115 base pairs in length and contains the octamer sequence ATGCAAAT in the top strand. The protection patterns shown are for the bottom strand of the probe. The autoradiogram shown represents sections of the same gel that has been rearranged for the clarity of presentation. The footprint of the POU homeodomain (lane 3) includes four of the five prominent bands of the intact POU domain (lane 2). The region of the DNA protected from DNase I cleavage by a mixture containing the POU homeodomain and the POU-specific domain corresponds precisely to the region protected by the intact POU domain (lane 4). Note that the addition of the POU-specific domain protects one prominent band (asterisk, lane 4) that is not protected by the isolated homeodomain. The protein concentrations in the binding reactions are: POU domain: 1×10^{-7} M; POU homeodomain: 2.5×10^{-6} M; POU-specific domain: 1×10^{-5} M.

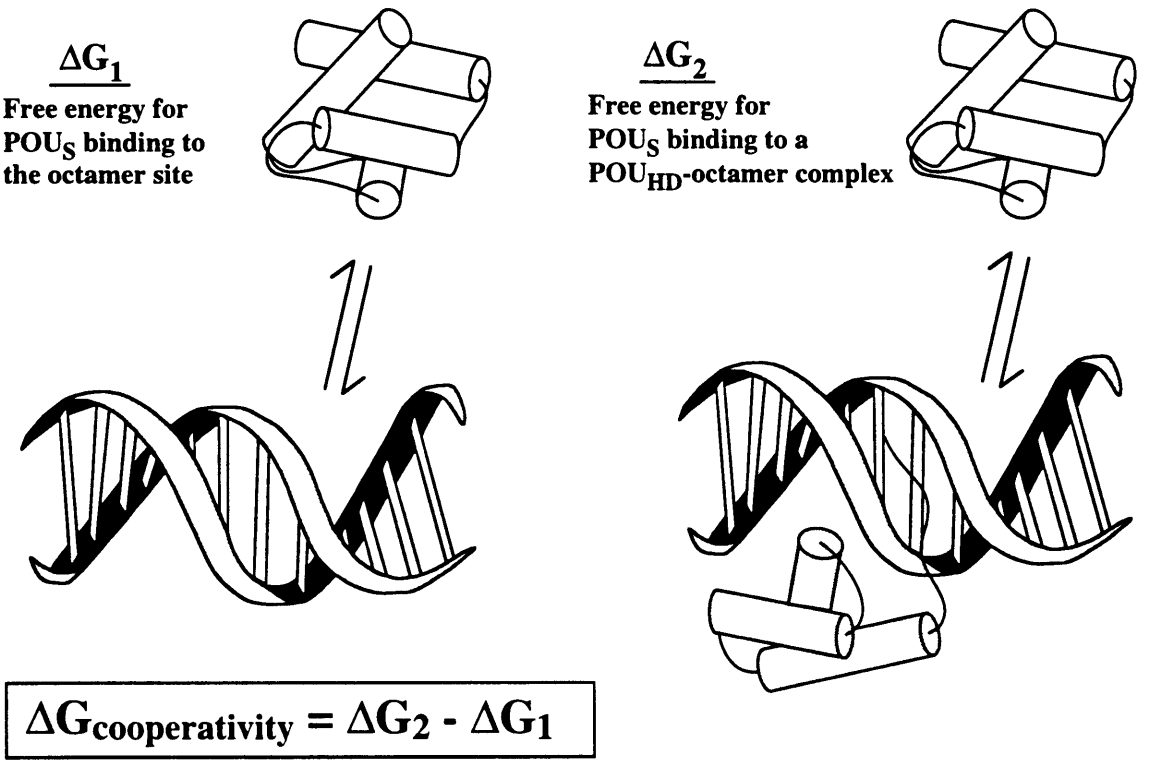

Figure 3. Sketch summarizing the strategy used to measure cooperative DNA-binding by the POU-specific domain and the POU homeodomain. The sketch on the left depicts the POU-specific domain binding to the octamer site, and this reaction has a free energy ΔG_1. The sketch on the right depicts the POU-specific domain binding to a POU-homeodomain—octamer complex, and this reaction has a free energy ΔG_2. [As described in the legend to Figure 4, binding was monitored by quantitative DNase I footprinting experiments.] The difference in the free energies of these reactions represents the free energy of cooperativity ($\Delta G_{cooperativity}$) for the POU-specific domain and the POU homeodomain.

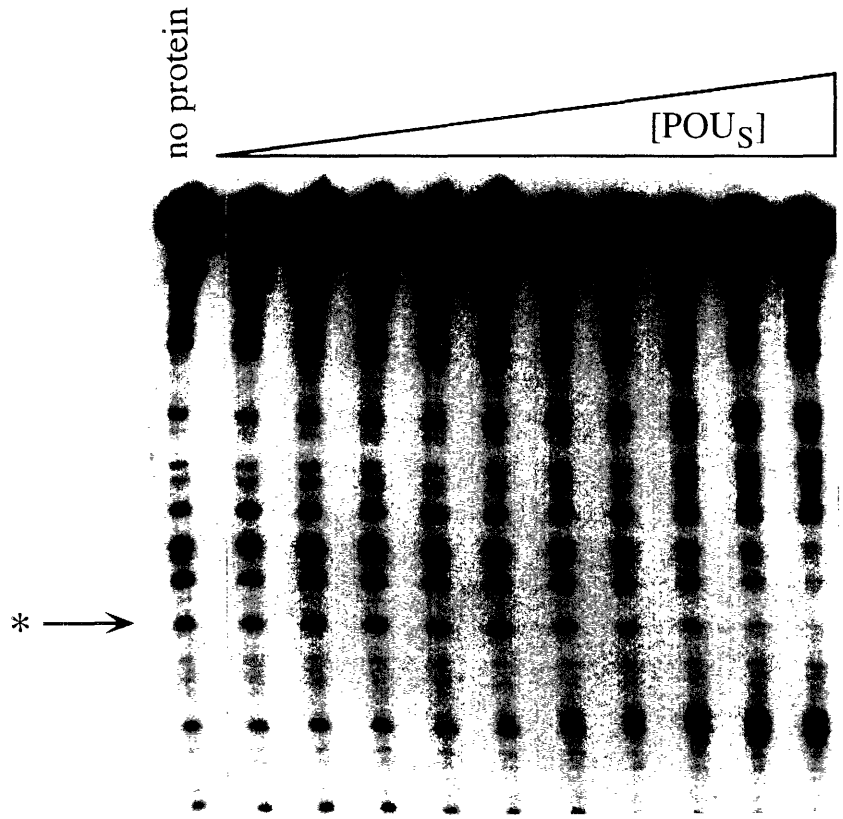

Figure 4. Quantitative DNase I footprinting experiment measuring the binding of the POU-specific domain to the octamer site. The concentration of the POU-specific domain was increased in two-fold increments from 9.76 x 10^{-8} M to 2 x 10^{-4} M. Disappearance of the band (asterisk) identified as specifically protected by the POU-specific domain (see legend to Figure 1) was quantitated as a function of the concentration of this domain.

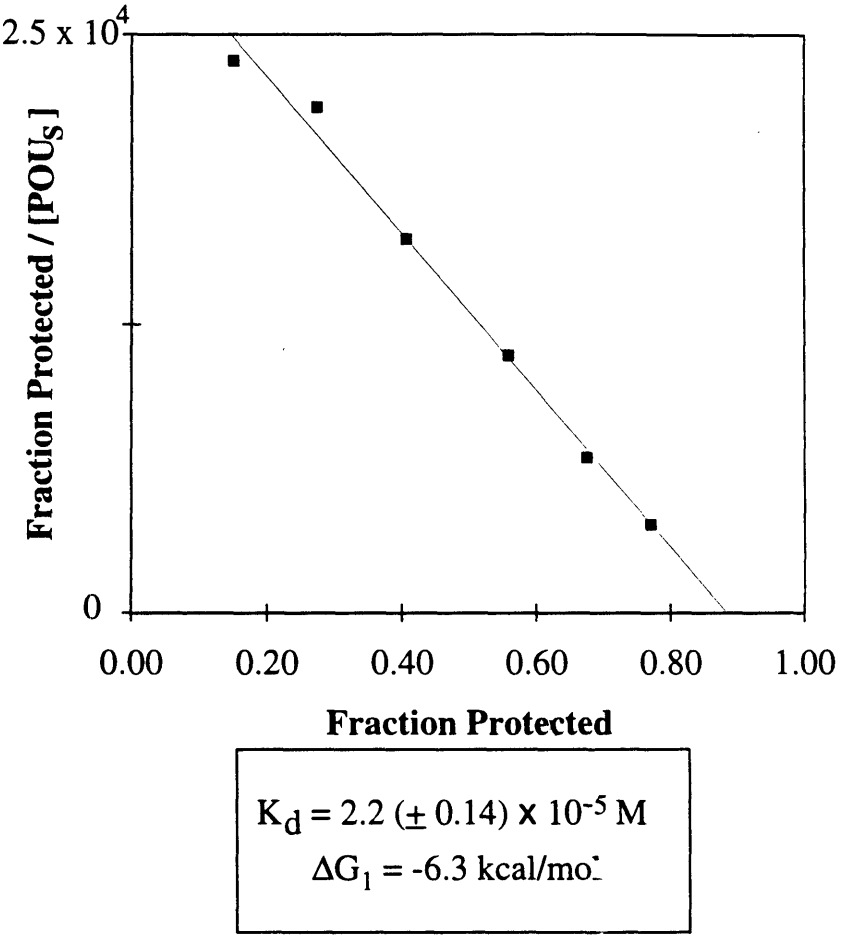

Figure 5. Scatchard plot for binding of the POU-specific domain to the octamer site.

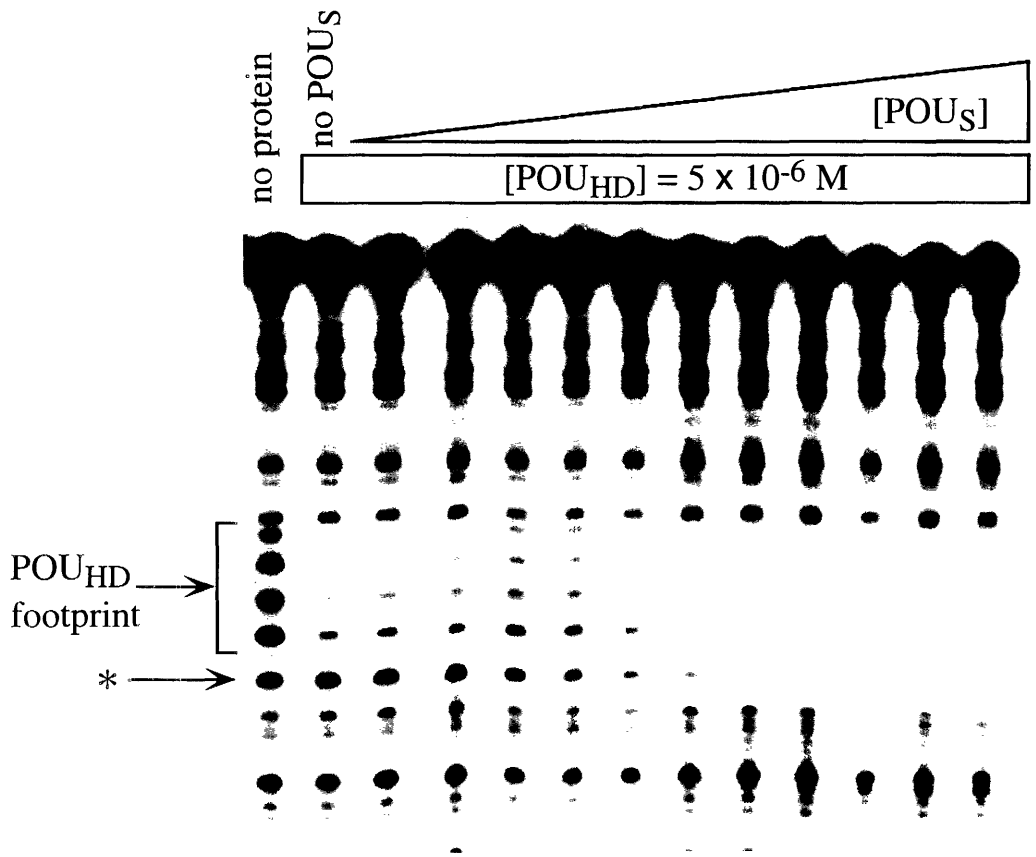

Figure 6. Quantitative DNase I footprinting experiment measuring binding of the POU-specific domain to a POU homeodomain—octamer complex. The POU homeodomain concentration is 5 x 10^{-6} M in all reactions, which should be sufficient to give an occupancy of 97%. The concentration of the POU-specific domain was increased in two-fold increments from 9.76 x 10^{-8} M to 2 x 10^{-4} M. Disappearance of the band (asterisk) that is specifically protected by the POU-specific domain was quantitated as a function the concentration of this domain.

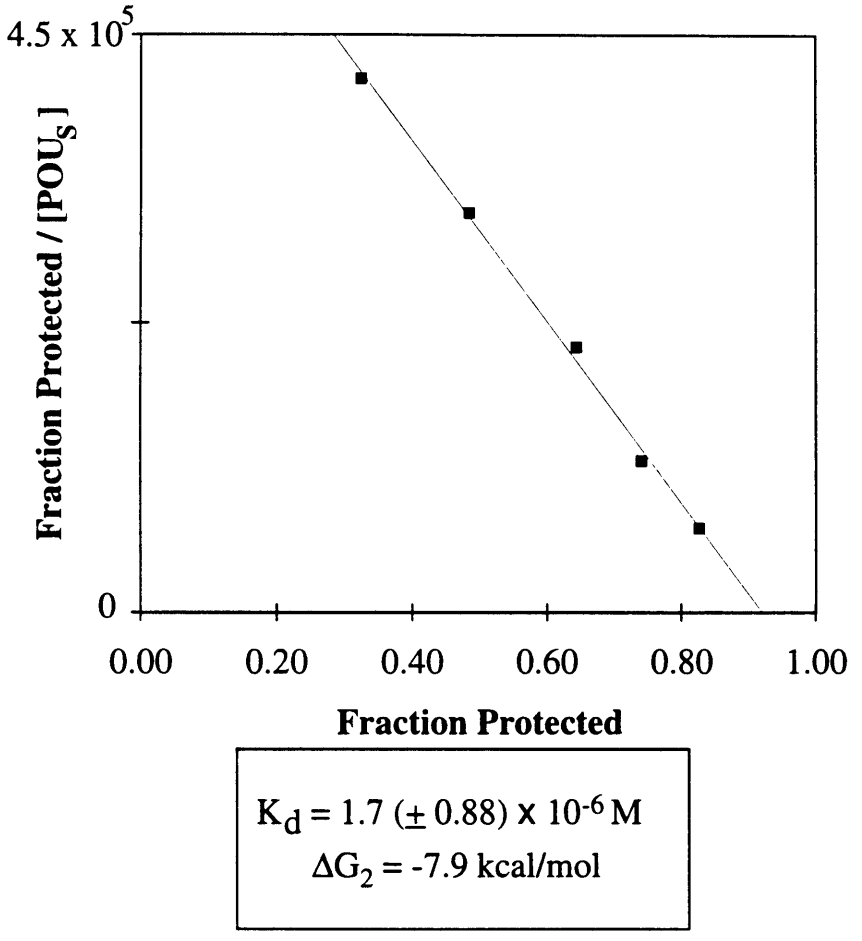

Figure 7. Scatchard plot for binding of the POU-specific domain to the POU homeodomain—octamer complex.

CHAPTER 4

DNA Recognition by the POU Domain: Outlook for Future Experiments

Introduction

Our structural and biochemical studies of Oct-1 POU domain—DNA interactions have provided a great deal of information about how its covalently linked subdomains are used in DNA recognition. Our crystallographic studies have shown that: 1) The POU-specific domain and the POU homeodomain do not contact one another when complexed with an octamer DNA site. 2) The subdomains makes specific base contacts to subsites that overlap at the center of the octamer - the POU-specific domain contacts the ATGCAAAT subsite and the POU homeodomain contacts the ATGCAAAT subsite. 3) The structure of the POU-specific domain is very similar to the structure of the λ repressor DNA-binding domain and it also docks on the DNA in the same way. Similarly, the structure of the POU homeodomain is nearly identical to the structure of other homeodomains and its interaction with DNA is also conserved. 4) The linker connecting the two subdomains is apparently very flexible, as it cannot be seen in the electron density maps.

Features of the crystal structure raised questions about coupling between the Oct-1 POU-specific domain and POU homeodomain and about the role of the covalent linker connecting these subdomains. To address these questions, we performed biochemical studies with the Oct-1 POU domain and its individual subdomains. These experiments have shown that: 1) The Oct-1 POU domain strongly prefers to bind the wild type octamer site. Inserting two base pairs near the center of the octamer site reduces the affinity 30-fold. 2) Individual polypeptides corresponding to the Oct-1 POU-specific domain and POU homeodomain bind the octamer sequence cooperatively, in the absence of the linker polypeptide. This cooperativity is likely to dictate

the preferred binding of Oct-1 to overlapping octamer subsites. 3) The chelate effect provided by the covalent linker also contributes to cooperative binding by increasing the effective concentration of the two subdomains in the context of the intact POU domain.

While our structural and biochemical studies have answered many questions regarding POU domain—DNA interactions, some questions remain unanswered and some new questions have been raised. Therefore, this final chapter addresses issues for future research regarding the POU family of transcription factors. The possible directions discussed include continued structural studies of POU domain proteins and further studies of subdomain cooperativity in POU domain—DNA interactions. Our studies of cooperativity in Oct-1 POU domain—DNA interactions have suggested that analogous studies of zinc finger—DNA interactions may be informative and these possibilities are discussed as well.

Continued structural studies of POU domain proteins

Continued structural studies of POU domain proteins might involve the following topics: 1) A higher resolution structure of an Oct-1 POU domain—octamer complex. 2) Crystallization of the Oct-1 POU domain with a DNA site on which it dimerizes. 3) Structural studies of other POU domain proteins. 4) Structural studies of the Oct-1 POU domain with interacting proteins. Each of these topics is discussed below.

A higher resolution structure of the Oct-1 POU domain—octamer complex
The crystal structure of the Oct-1 POU domain—octamer complex was solved at 3 Å resolution and our high quality MIR map allowed us to trace the

protein chains unambiguously and to position the vast majority of side chains confidently. However, a higher resolution structure of the complex would be desirable for many reasons: 1) Fundamentally, the positions of the atoms for the protein and for the DNA would be determined more precisely. 2) Ordered water molecules could be included in the model. In particular, water-mediated protein-DNA interactions may become apparent at higher resolution. 3) Structural preferences for poorly ordered side chains and perhaps even for regions of the linker may become apparent with higher resolution data.

One strategy for obtaining a higher resolution structure would be to collect diffraction data on flash-frozen crystals. In many cases, the highest resolution to which accurate data can be recorded increases with frozen crystals, either because of a reduction in thermal disorder or simply as a result of reduced radiation damage and longer exposure times (for a review of cryocrystallography, see Rodgers, 1994). Although our past attempts to find a suitable cryoprotectant for the Oct-1—DNA complex crystals were unsuccessful, it may be worthwhile to continue these pursuits, as there is now more information and experience available in the lab for doing cryocrystallography. In addition, a mirror system recently installed on the x-ray system in the Pabo lab increases the intensity of the x-ray beam at least 4-fold, which may also allow higher resolution data to be collected, even in the absence of flash-freezing.

A very different approach towards a higher resolution structure of an Oct-1—octamer complex would be to eliminate the floppy 24 amino acid linker connecting the subdomains. It is possible that these flexible residues contribute to disorder in the crystal lattice, limiting the extent to which these crystals diffract. Our studies with the individual POU-specific domain and

POU homeodomain polypeptides (Chapter 3) show that they can form a POU domain-like complex with the octamer sequence, in the absence of any linker residues. Thus, a ternary complex of the two individual subdomains bound to the octamer sequence may form more highly ordered crystals than the binary complex of the POU domain and the octamer sequence. Additionally, it would be interesting to see whether the subdomains are positioned on the DNA in precisely the same way in the absence of the linker as they are in the presence of the linker. Crystallization of each subdomain individually with DNA containing the octamer site could also be informative. This approach may yield crystals that diffract to higher resolution than crystals with the intact POU domain, allowing a more detailed analysis of the DNA contacts made by the POU-specific domain and the POU homeodomain.

Crystallization of Oct-1 with an octamer-heptamer site

Although Oct-1 binds as a monomer to the octamer site, there are a class of binding sites to which Oct-1 binds cooperatively as a dimer, and it would be very interesting to study such a complex crystallographically. In addition to an octamer element, immunoglobulin heavy-chain genes contain a heptamer sequence (CTCATGA), located 2, 14, or 22 base pairs upstream of the octamer sequence (Kemler et al., 1989; Poellinger et al., 1989). Although the heptamer sequence is not closely related to the octamer sequence (ATGCAAAT), it is recognized by Oct-1 and Oct-2, (with low affinity) and binding is strongly facilitated by occupancy of the adjacent octamer site (Kemler et al., 1989; LeBowitz et al., 1989; Poellinger and Roeder, 1989; Poellinger et al., 1989). The POU domain itself is sufficient for this cooperative interaction (LeBowitz et al., 1989; Verrijzer et al., 1992). A structure of the Oct-1 POU domain bound to an octamer-heptamer composite site would reveal not only the mechanism

with which two POU domains can bind cooperatively to DNA, but also how the Oct-1 POU domain interacts with a DNA sequence distinct from the octamer sequence.

Structural studies of other POU domain proteins

While the structure of the Oct-1 POU domain—octamer complex has provided us with a higher level of understanding of POU domain—DNA interactions, there may be fundamental differences in how other POU domains interact with DNA. It will be important to pursue structural studies of other POU domains complexed with DNA to expand our understanding of DNA recognition by the POU family of transcription factors.

One attractive candidate for structural studies is the POU domain protein Brn-2 because biochemical studies suggest that this protein (and others in its subclass) interacts with DNA in a manner distinct from Oct-1. Li et al. (1993) have shown that this POU domain prefers a DNA site in which the orientation of the POU-specific subsite is reversed relative to its orientation in the octamer sequence (ATGCA -> TGCAT). Additionally, Brn-2 prefers to bind a site in which the subsites are separated by 2 base pairs, but it can also bind subsites with spacings of 0 or 3 base pairs. Ideally, one would like to compare the structures of Brn-2 bound to DNA sites with all three spacings. Such a study would yield a great deal of insight into the different modes of DNA recognition POU domain proteins can adopt.

The Pit-1 POU domain is also a compelling choice for continued structural studies of this DNA-binding motif. Studies of Pit-1 would not only provide additional information about POU domain—DNA interactions, but they may also be of medical interest. Pit-1 controls expression of pituitary-specific genes, and point mutations in both the POU-specific domain and the

POU homeodomain of Pit-1 are associated with human dwarfism (Radovick et al., 1992; Pfäffle et al., 1992). DNA-binding studies suggest that Pit-1 binds its recognition element as a dimer, and this dimerization may be mediated by the POU-specific domain (Ingraham et al., 1990). It is not clear from the binding site used in these studies (TTC**ATGAATAT**ATATATAATCAGG; Pit-1 consensus in bold) how the "second" POU domain in this complex might bind. [The "first" Pit-1 POU domain apparently binds with its POU-specific domain positioned over the ATGAA portion of the site and its POU homeodomain positioned over the ATAT portion of the site (Ingraham et al., 1990).] Thus, a crystal structure of a Pit-1 POU domain—DNA complex would reveal how these POU domains can dimerize on DNA. An additional interesting feature of Pit-1 is that the linker connecting the subdomains is nine residues shorter than the Oct-1 linker. This shorter linker spanning the distance between the POU-specific domain and the POU homeodomain may be less flexible and perhaps could be seen in electron density maps.

Crystallization of the Oct-1 POU domain with interacting proteins

Although the activation of transcription generally involves the formation of multiprotein complexes on DNA, most structural studies to date have focused on the interaction of a single DNA-binding protein with DNA. An important next step in structural studies of transcription factors will be to examine interactions between heterologous proteins bound to DNA. Oct-1 has been shown to interact with a number of different proteins when bound to DNA, often through its POU domain. Thus, the Oct-1 POU domain may provide a useful model system to begin understanding higher order protein-DNA complexes at a structural level.

The interaction between Oct-1 and the herpes simplex virus protein

VP16 (α-TIF) has been extensively characterized physiologically and biochemically; a crystallographic study would allow these data to be interpreted in the context of a three-dimensional structure. VP16 activates expression of the virus immediate early genes by binding with Oct-1 and another host cell factor (HCF) at the DNA target sequence TAATGARAT (Gerster and Roeder, 1988; Kristie et al., 1989; Stern et al., 1989). Formation of the complex between VP16 and Oct-1 is dependent on specific amino acids on the surface of the Oct-1 homeodomain (Stern et al., 1989; Pomerantz et al., 1992; Lai et al., 1992) and in fact, the POU homeodomain alone is sufficient for complex formation (Kristie and Sharp, 1990). A structural study of the Oct-1 POU domain complexed with VP16 at a TAATGARAT site would provide important information about the nature of this complex, including: 1) the interactions between these proteins when bound to DNA; 2) whether VP16 makes specific DNA contacts; and 3) how the POU domain interacts with this site.

Structural studies of a ternary complex containing the Oct-1 POU domain and the B cell-specific protein Bob-1/OBF-1 bound to the octamer sequence would provide a great deal of insight into coactivator-transcription factor interactions. Bob-1/OBF-1 was isolated by two groups who were in search of factors that interact with Oct-1 or Oct-2 to enhance activation of lymphoid-specific genes (Gstaiger et al., 1995; Strubin et al., 1995). This proline-rich protein contains 256 amino acids and by itself has no apparent DNA-binding activity, but it forms a specific complex with the POU domains of Oct-1 and Oct-2. A chimeric protein containing Bob-1/OBF-1 fused to the GAL4 DNA-binding domain is able to activate transcription from a reporter plasmid containing a GAL4 recognition element, suggesting that this protein can function as a transcriptional activator (Gstaiger et al., 1995). Additionally,

expression of this protein in non-lymphoid cells was shown to stimulate expression of an immunoglobulin light chain promoter, presumably through interactions with the ubiquitously expressed Oct-1 (Strubin et al., 1995). Thus, the structure of a complex containing the Oct-1 POU domain and Bob-1/OBF-1 bound to an octamer site would be an important step toward understanding transcription factor-coactivator interactions.

Studies of subdomain cooperativity and linker composition in POU domain—DNA interactions

Our biochemical studies have shown that individual peptides corresponding to the Oct-1 POU-specific domain and POU homeodomain bind cooperatively to the octamer sequence, in the absence of the linker polypeptide. Further studies of this effect may involve: 1) measuring binding of the Oct-1 POU domain to an octamer site with a one base pair insertion at the center of the octamer site; 2) identification of the residues that mediate this cooperative binding; and 3) investigation of subdomain coupling in other POU domain proteins.

Measuring Oct-1 POU domain binding to subsites separated by one base pair
We have shown that binding of the Oct-1 POU domain to an octamer site with two base pairs inserted near the center of the site (ATGCACAAAT) is reduced by 2.1 kcal/mol and we have proposed that much of this loss may be attributed to the lack of coupling between the subdomains on this site (which contributes 1.6 kcal/mol to POU domain binding). Would there be a similar reduction in affinity if one base pair were added near the center of the octamer site? The site to use for such a measurement should be

ATGCAAAAT, which should maintain all specific base contacts by both subdomains. Thus, any reduction in affinity would most likely be due to the loss of coupling between the subdomains.

Identification of residues involved in cooperative binding

In Chapter 3, we proposed that cooperative binding by the Oct-1 POU-specific domain and POU homeodomain is likely to be mediated via overlapping DNA contacts near the center of the octamer site. To determine which residues are involved, one might eliminate some of these overlapping contacts and test whether cooperative binding was abolished. The A/T base pair at position 5 of the octamer is contacted by both the POU-specific domain and the POU homeodomain. Leu-55 of the POU-specific domain is in van der Waals' contact with the C5 methyl group of the thymine, and Arg-5 in the N-terminal arm of the POU homeodomain makes a hydrogen bond to the N3 of the adenine. Substituting a uracil for the thymine would eliminate the POU-specific domain interaction at this base pair, as uracil lacks a C5 methyl group. This may affect affinity, but a reduction or lack of cooperative binding on this site would indicate that coupling is mediated through overlapping contacts to this base pair.

Subdomain coupling in other POU domain—DNA interactions

Our studies with the isolated POU-specific domain and POU homeodomain from Oct-1 have shown that these subdomains bind DNA cooperatively in the absence of the linker polypeptide (Chapter 3). Is this effect specific to Oct-1 or is this a more general feature of POU domain—DNA interactions? A similar study of the Pit-1 POU-specific domain and POU homeodomain would reveal whether this mechanism was common to POU domains which

(presumably) bind in a similar orientation as Oct-1. [The Oct-2 POU domain is nearly identical to Oct-1 and it seems likely that it will behave identically.] It would also be of interest to measure coupling between the subdomains of proteins such as Brn-2 which 1) prefer to bind separated subsites and 2) bind with their POU-specific domains in the opposite orientation relative to the Oct-1—DNA complex (Li et al., 1993). There are currently no structures of these other POU domains complexed with DNA, so we do not know whether the POU-specific and POU homeodomain interact in the complex; however, it will be important to determine whether cooperative effects are involved in determining the preferred orientation and spacing of subdomains in these systems.

Importance of linker length and composition in POU domain-DNA interactions

In our crystal structure of the Oct-1 POU domain—octamer complex, the linker connecting the POU-specific domain and the POU homeodomain was not visible; therefore, we have assumed that it is highly flexible and unstructured. Biochemical evidence also suggests that the linker merely serves as a flexible tether between the subdomains (see Chapter 2). Nonetheless, it would be informative to directly address whether the linker can influence POU domain—DNA interactions.

By swapping linkers between POU domains, one can test whether the linker confers any DNA-binding preferences on these proteins. Aurora and Herr (1992) exchanged linkers between Oct-1 and Pit-1 and tested binding to physiologically relevant Oct-1 binding sites. For nearly all of the sites tested, replacing the Oct-1 linker (24 residues) with the Pit-1 linker (15 residues) significantly decreased binding by the Oct-1 POU domain. This

result shows that the linker does have some influence on DNA binding, but it is not clear whether this effect is due to the decrease in linker length (by nine residues) or whether it is due to the very different amino acid sequences of these linkers:

```
Oct-1:  NLSSDSSLSSPSALNSPGIEGLSR
Pit-1:  QVGALYNEKVGANER
```

In order to test what effect the amino acid composition of the linker has on DNA binding (in the absence of a significant difference in linker length), it would be interesting to substitute the Brn-5 linker (20 residues) (Andersen et al., 1993) for the Oct-1 linker, and measure DNA binding to the octamer sequence. These two linkers are very dissimilar in amino acid sequence and composition, but they are similar in length (4 residues difference):

```
Oct-1:  NLSSDSSLSSPSALNSPGIEGLSR
Brn-5:  LRNQEGQQNLMEFVGGEPSK
```

A decrease in affinity of the chimeric Oct-1 protein for the octamer site would suggest that the sequence of the linker does influence DNA recognition.

It would also be very interesting to replace the Oct-1 linker with the much longer linker from the CEH-18 POU domain (56 amino acids) (Greenstein et al., 1994):

```
Oct-1:  NLSSDSSLSSPSALNSPGIEGLSR
CEH-18: MAIEGGATVTDLIDKKTIHNGNHHTIHHVDIHETSISNSISS
        VTASSLLSREQHVK
```

In the absence of any other effect, inserting a significantly longer linker into the Oct-1 POU domain would be expected to increase its dissociation constant by decreasing the effective concentration provided by the linker.

Studies of cooperativity in zinc finger-DNA interactions

Although TFIIIA-type zinc finger proteins have been studied in great detail, our studies of Oct-1 POU domain—DNA interactions suggest new approaches toward understanding how these linked modules cooperate in DNA recognition. It would be very interesting to extend our studies of coupling and the chelate effect in Oct-1 POU domain—DNA interactions to the analysis of these effects in TFIIIA-type zinc finger—DNA interactions. As discussed in Chapter 1, zinc fingers are comprised of tandemly repeated finger modules connected by short, conserved linkers. Is there coupling between adjacent fingers that is analogous to the coupling between the Oct-1 POU-specific domain and POU homeodomain? What is the contribution of the linker to DNA binding?

Features seen in the zinc finger—DNA crystal structures strongly suggest that binding by adjacent fingers may be coupled in a manner analogous to the coupling between the Oct-1 POU subdomains. Zinc finger domains often make DNA contacts in the subsites of adjacent fingers. Of the three zinc finger—DNA complexes for which there are crystal structures, GLI may be the most appealing prospect for studying finger-finger coupling. Fingers 4 and 5 make extensive overlapping phosphate contacts (Pavletich and Pabo, 1993) and it would be interesting to test whether a "severed" finger 5 could be added "in trans" to bind cooperatively with fingers 1 - 4.

Further studies directly characterizing the properties of the linkers between zinc fingers would also be informative. As described in Chapter 1, experiments by Choo and Klug (1993) using the first three fingers of TFIIIA suggest that the conserved linker residues - TGEK - are important for optimal DNA binding. Additionally, they showed that inserting a single serine residue between the G and E or between the E and K resulted in about 10-fold

weaker binding to DNA. Would binding be affected even more drastically if longer insertions were made in this region, or would the effects be less than that of adding the single residue? The linker between the two zinc fingers of the Tramtrack protein diverges significantly from the TGEK consensus: it is one residue longer and has the sequence KRNVK. Is Tramtrack as sensitive to substitutions in its linker as TFIIIA? Could a TGEK linker be substituted for the KRNVK linker? Understanding the requirements and restrictions for zinc finger linkers will be particularly useful for linking together novel combinations of zinc fingers with new DNA-binding specificities.

Conclusions

Our studies of the Oct-1 POU domain have allowed us to describe in both structural and energetic terms how the POU-specific domain and the POU homeodomain cooperate to recognize the octamer site. We have shown that these subdomains are structurally independent but they recognize neighboring, partially overlapping subsites. We have demonstrated that binding of these subdomains is coupled through covalent linkage and through overlapping DNA contacts. These observations may be of general importance for understanding how other DNA-binding domains of covalently-linked modules interact with DNA. Key issues involve: 1) interactions between subdomains; 2) the properties of the linker connecting the modules; and 3) the contributions of individual subdomains to DNA binding. Covalently-linked DNA-binding modules provide a new level of complexity in understanding protein-DNA interactions, and our characterization of the Oct-1 POU domain has provided a comprehensive description of one such system.

References

Andersen, B., Schonemann, M. D., Pearse, R. V. 2d, Jenne, K., Sugarman, J., and Rosenfeld, M. G. (1993). Brn-5 is a divergent POU domain factor highly expressed in layer IV of the neocortex. J. Biol. Chem. *268*, 23390-23398.

Aurora, R., and Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. Mol. Cell. Biol. *12*, 455-467.

Choo, Y. and Klug, A. (1993). A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. Nuc. Acids Res. *21*, 3341-3346.

Gerster, T. and Roeder, R. G. (1988). A herpes virus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. Prot. Natl. Acad. Sci. USA *85*, 6347-6351.

Greenstein, D., Hird, S., Plasterk, R. H. A., Andachi, Y., Kohara, Y., Wang, B., Finney, M., and Ruvkin, G. (1994). Targeted mutations in the Caenorhabditis elegans POU homeo box gene ceh-18 cause defects in oocyte cell cycle arrest, gonad migration, and epidermal differentiation. Genes & Dev. *8*, 1935-1948.

Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W., and Hovens, C. M. (1995). A B-cell coactivator of octamer-binding transcription factors. Nature *373*, 360-362.

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1—Pit-1 interactions. Cell *61*, 1021-1033.

Kemler, I., Schreiber, E., Müller, M. M., Matthias, P., and Schaffner, W. (1989). Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. EMBO J. *8*, 2001-2008.

Kristie, T. M., LeBowitz, J. H., and Sharp, P. A. (1989). The octamer-binding proteins form multi-protein-DNA complexes with the HSV α-TIF regulatory protein. EMBO J. *8*, 4229-4238.

Kristie, T. M., and Sharp, P. A. (1990). Interactions of the Oct-1 POU subdomains with specific DNA sequences and the HSV α-*trans*-activator protein. Genes & Dev. *4*, 2383-2396.

Lai, J -S. Cleary, M. A., and Herr, W. (1992). A single amino acid exchange transfers VP16-induced positive control from the Oct-1 to the Oct-2 homeo domain. Genes & Dev. *6*, 2058-2065.

LeBowitz, J. H., Clerc, R. G., Brenowitz, M., and Sharp, P. A. (1989). The Oct-2

protein binds cooperatively to adjacent octamer sites. Genes & Dev. *2*, 1227-1237.

Li, P., He, X., Gerrero, M. R., Mok, M., Aggarwal, A., and Rosenfeld, M. G. (1993). Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. Genes & Dev. *7*, 2483-2496.

Pavletich, N. P., and Pabo, C. O. (1993). Crystal structure of a five finger GLI-DNA complex: new perspectives on zinc fingers. Science *261*, 1701-1707.

Pfäffle, R. W., DiMattia, G. E., Parks, J. S., Brown, M. R., Wit, J. M., Jansen, M., Van der Nat, H., Van den Brande, J. L., Rosenfeld, M. G., and Ingraham, H. A. (1992). Mutation of the POU-specific domain of Pit-1 and hypopituitarism without hypoplasia. Science *257*, 1118-1121.

Poellinger, L., and Roeder, R. G. (1989). Octamer transcription factors 1 and 2 each bind to two different functional elements in the immunoglobulin heavy-chain promoter. Mol. Cell. Biol. *9*, 747-756.

Poellinger, L. Yoza, B. K., Roeder, R. G. (1989). Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter. Nature *337*, 573-576.

Pomerantz, J. L., Kristie, T. M., and Sharp, P. A. (1992). Recognition of the surface of a homeo domain protein. Genes & Dev. *6*, 2047-2057.

Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995). Structure-based design of transcription factors. Science *267*, 93-96.

Radovick, S., Nations, M., Du, Y., Berg, L. A., Weintraub, B. D., and Wondisford, F. E. (1992). A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. Science *257*, 1115-1118.

Rodgers, D. W. (1994). Cryocrystallography. Structure *2*, 1135-1140.

Stern, S., Tanaka, M., and Herr, W. (1989). The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature *341*, 624-630.

Strubin, M., Newell, J. W., and Matthias, P. (1995). OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. Cell *80*, 497-506.

Verrijzer, C. P., van Oosterhout, J. A. W. M., and van der Vliet, P. C. (1992). The Oct-1 POU domain mediates interactions between Oct-1 and other POU proteins. Mol. Cell. Biol. *12*, 542-551.